

5-6-2014

Advanced Methodologies in Dynamic Traffic Assignment Modeling of Managed Lanes

Shaghayegh Shabanian
sshab001@fiu.edu

DOI: 10.25148/etd.FI14040896

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>

 Part of the [Civil and Environmental Engineering Commons](#)

Recommended Citation

Shabanian, Shaghayegh, "Advanced Methodologies in Dynamic Traffic Assignment Modeling of Managed Lanes" (2014). *FIU Electronic Theses and Dissertations*. 1238.
<https://digitalcommons.fiu.edu/etd/1238>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

ADVANCED METHODOLOGIES IN DYNAMIC TRAFFIC ASSIGNMENT
MODELING OF MANAGED LANES

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CIVIL ENGINEERING

by

Shaghayegh Shabani

2014

To: Dean Amir Mirmiran
College of Engineering and Computing

This dissertation, written by Shaghayegh Shabanian, and entitled Advanced Methodologies in Dynamic Traffic Assignment Modeling of Managed Lanes, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Albert Gan

Xia Jin

David Shen

Zhenmin Chen

Mohammed Hadi, Major Professor

Date of Defense: March 6, 2014

The dissertation of Shaghayegh Shabanian is approved.

Dean Amir Mirmiran
College of Engineering and Computing

Dean Lakshmi N. Reddi
University Graduate School

Florida International University, 2014

DEDICATION

I dedicate this dissertation to my beloved parents, Farkhonde and Mohammad, for their unconditional love and their endless support and encouragement.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Dr. Mohammed Hadi, for his guidance, patience, and for his confidence in me. It was my honor to work with such a well-respected researcher in the Intelligent Transportation Systems (ITS) field. He was generous with the time he spent on advising me and training me to conduct high quality research, and to develop a variety of skills that are necessary to succeed in my professional career.

I experienced a great teamwork environment in the Integrated Intelligent Transportation Systems (IITS) research lab. I would like to thank Dr. Yan Xiao, who is incredibly wise, knowledgeable, and kind to share her knowledge. I appreciate her continued support of me in different endeavors, including leading student chapters at Florida International University. I would like extend my deep appreciation to Dr. Halit Ozen, who I learned much from in different transportation engineering topics, especially planning and modeling. I would also like to thank my friend and colleague, Dr. Patricio Alvarez, for his valuable advice on professional and academic development.

I am also grateful to my committee members for serving on my committee, and for their valuable comments, suggestions, and interest in my research.

I will be always thankful to my dear friends Nahid and Somaye for their assistance and friendship, and to my patient husband, James, for his continued support, care and encouragement.

ABSTRACT OF THE DISSERTATION
ADVANCED METHODOLOGIES IN DYNAMIC TRAFFIC ASSIGNMENT
MODELING OF MANAGED LANES

by

Shaghayegh Shabani

Florida International University, 2014

Miami, Florida

Professor Mohammed Hadi, Major Professor

Managed lane strategies are innovative road operation schemes for addressing congestion problems. These strategies operate a lane (lanes) adjacent to a freeway that provides congestion-free trips to eligible users, such as transit or toll-payers. To ensure the successful implementation of managed lanes, the demand on these lanes need to be accurately estimated. Among different approaches for predicting this demand, the four-step demand forecasting process is most common. Managed lane demand is usually estimated at the assignment step. Therefore, the key to reliably estimating the demand is the utilization of effective assignment modeling processes.

Managed lanes are particularly effective when the road is functioning at near-capacity. Therefore, capturing variations in demand and network attributes and performance is crucial for their modeling, monitoring and operation. As a result, traditional modeling approaches, such as those used in static traffic assignment of demand forecasting models, fail to correctly predict the managed lane demand and the associated system performance. The present study demonstrates the power of the more

advanced modeling approach of dynamic traffic assignment (DTA), as well as the shortcomings of conventional approaches, when used to model managed lanes in congested environments. In addition, the study develops processes to support an effective utilization of DTA to model managed lane operations.

Static and dynamic traffic assignments consist of demand, network, and route choice model components that need to be calibrated. These components interact with each other, and an iterative method for calibrating them is needed. In this study, an effective standalone framework that combines static demand estimation and dynamic traffic assignment has been developed to replicate real-world traffic conditions.

With advances in traffic surveillance technologies collecting, archiving, and analyzing traffic data is becoming more accessible and affordable. The present study shows how data from multiple sources can be integrated, validated, and best used in different stages of modeling and calibration of managed lanes. Extensive and careful processing of demand, traffic, and toll data, as well as proper definition of performance measures, result in a calibrated and stable model, which closely replicates real-world congestion patterns, and can reasonably respond to perturbations in network and demand properties.

TABLE OF CONTENTS

CHAPTER	PAGE
1. INTRODUCTION.....	13
1.1. Background	13
1.2. Problem Statement	16
1.3. Research Goal and Objectives.....	19
1.4. Dissertation Organization.....	19
2. LITERATURE REVIEW.....	22
2.1. Managed Lane Modeling Frameworks	22
2.2. Dynamic Traffic Assignment.....	30
2.3. Supply/Network Calibration	34
2.3.1. Mesoscopic Simulation Supply Calibration.....	35
2.3.2. Bottleneck Identification.....	37
2.3.3. Free-Flow Speed	38
2.3.4. Capacity Definition and Estimation.....	39
2.3.5. Traffic Flow Model Parameter Estimation	42
2.4. Demand Estimation.....	45
2.5. Joint Supply and Demand Calibration	52
2.6. Convergence	54
2.7. Summary.....	58
3. METHODOLOGY	61
4. DATA ACQUISITION AND VALIDATION.....	66
4.1. Network and Demand Data Extraction	66
4.1.1. Subarea Network and Matrix Extraction	66
4.1.2. Demand Data	70
4.2. Detector Data Acquisition and Preprocessing	70
4.3. Other Data Sources	72
4.4. Data Preprocessing and Validation.....	72
4.5. Summary.....	79
5. SUPPLY CALIBRATION	81
5.1. Bottleneck Identification.....	83
5.2. Free-Flow Speed	83
5.3. Capacity Estimation.....	85
5.4. Coded Capacity Impacts	89
5.5. Other Traffic Flow Model Parameter Estimation.....	93
5.6. Summary.....	98
6. ORIGIN-DESTINATION MATRIX ESTIMATION.....	100
6.1. Static OD Estimation	103
6.2. Dynamic Matrix Estimation.....	111
6.3. Performance Measures and Matrix Adjustment	113
6.4. Recommendations for OD Estimation Improvement	115

6. 5. Summary	116
7. TRAFFIC ASSIGNMENT	118
7.1. Route Choice Behavior Modeling	118
7.2. Derivations from Observed Data	121
7.3. Calibrating the Value of Time	125
7.4. Calibrating the Willingness-to-Pay Parameters	127
7.5. Comparing Static and Dynamic Traffic Assignment.....	128
7.6. Convergence and Stability	130
7.7. Model Validation and Sensitivity Analysis	132
7.8. Assessment of Performance Measures.....	135
7.9. Summary	140
8. CONCLUSION AND FUTURE STUDY	142
REFERENCES	148
VITA	157

LIST OF TABLES

TABLE	PAGE
Table 4-1 Available and Selected Detectors	71
Table 5-1 Estimated Capacity at Active Bottleneck Locations (VPH).....	87
Table 5-2 Goodness-of-Fit Measures.....	96
Table 7-1 Initial Not-Willing-to-Pay Proportion for Cost per Time Saved by Demand Category	120
Table 7-2 Default Toll Values Based on the ML Maximum Density.....	121
Table 7-3 Implemented Toll Value for I-95 Northbound ...	Error! Bookmark not defined.
Table 7-4 Implemented Toll Value for I-95 Northbound	126
Table 7-5 Goodness of Fit Statistics for Volume Replication	138

LIST OF FIGURES

FIGURE	PAGE
Figure 2-1 Interaction Between Demand and Supply.....	50
Figure 3-1. Methodology Flowchart.....	62
Figure 4-1 The Extracted Subarea from the SERPM Model.....	67
Figure 4-2 Network Geometry and Distance Update Procedure Flow Charts.....	68
Figure 4-3 Network Curvature Correction.....	69
Figure 4-4 Volume Inconsistency Between Successive Detectors.....	74
Figure 4-5 Volume Inconsistency Between Successive Detectors.....	75
Figure 4-6 Comparison of PTMS vs ITS Volume Data (August 9, 2011).....	76
Figure 4-7 Comparison of PTMS vs ITS Volume Data (August 10, 2011).....	77
Figure 4-8 Comparison of PTMS vs ITS Volume Data (August 11, 2011).....	77
Figure 4-9 Lane by Lane Speed Data.....	78
Figure 4-10 Lane by Lane Occupancy Data.....	79
Figure 5-1 Speed Contour for Different Classes of Demand.....	82
Figure 5-2 Variation of the FFS Along the Corridor (I-95 NB).....	84
Figure 5-3 Rakha Model Fitting for Capacity and TFM Parameters Estimation.....	88
Figure 5-4 Demonstrating breakdown flow and queue discharge rate.....	88
Figure 5-5 Speed Contour Maps for Static and Dynamic Traffic Assignment with Different Capacity Values.....	90
Figure 5-6 Travel Time Difference Between GPL and ML.....	91
Figure 5-7 Percentage of Travelers Diverting to ML.....	92
Figure 5-8 Effect of Implementing Different TFMs on Travel Speed.....	94
Figure 5-9 Difference Between Storage and Queuing Density on Travel Speed.....	95
Figure 6-1. Effect of Incorporating Queue in OD Estimation.....	108

Figure 6-2 Temporal Profile of Initial, Analyst, and Analyst Drive OD	109
Figure 6-3 Screenline Volume Replication by Factored Regional Matrix	109
Figure 6-4 Screenline Volume Replication by Analyst	110
Figure 6-5 Screenline Volume Replication by Analyst Drive	110
Figure 6-6 Comparing Initial and Estimated OD for One Pair	111
Figure 7-1 Managed Lane Model Based on Willingness-to-Pay Curve	119
Figure 7-2 Comparison Between Implemented and Calculated Toll for a Lightly-Congested Day	122
Figure 7-3 Comparison Between Implemented and Calculated Toll for Highly-Congested Intervals	123
Figure 7-4 Default and Developed Toll Density Curve	124
Figure 7-5 Comparison of Diverted Volume to ML for Different Toll Curves.....	125
Figure 7-6 Comparison of Diverted Volume to ML for Different VOTs.....	127
Figure 7-7 Different Shapes of Willingness-to-Pay Curve.....	128
Figure 7-8 Diverted Volume to ML Associated with Willingness-to-Pay Curves.....	128
Figure 7-9 Comparison Between Modeled and Observed ML Volume for Generalized Cost Function Method	129
Figure 7-10 Comparison Between Modeled and Observed ML Volume for Willingness-to-Pay Curve Method	130
Figure 7-11 Relative Gap for Different Assignment Approaches	131
Figure 7-12 Diverted Volume to ML for Different Assignment Approaches	131
Figure 7-13 Speed Contour for GPL with Different Demand Level in Generalized Cost Function Assignment.....	133
Figure 7-14 Speed Contour for GPL with Different Demand Level in Willingness-to Pay Curve Assignment	133
Figure 7-15 Changing in VMT and VHT with Changing Demand in Generalized Cost Function Assignment.....	134

Figure 7-16. Changing in VMT and VHT with Changing Demand in Willingness-to Pay Curve Assignment	134
Figure 7-17. Effect of Randomness in Simulation-Based DTA Tool.....	135
Figure 7-18 Scatter Plot of Observed vs. Simulated Volume of Screenlines	136
Figure 7-19 Scatter Plot of Observed Volume of Screenlines for Different Days	137
Figure 7-20 Comparison Between Observed and Simulated Flow on ML.....	137
Figure 7-21 Comparison Between Simulated and Observed Speed Contour	138
Figure 7-22 Scatter Plot of Observed vs. Simulated Speed of Screenlines	139
Figure 7-23 Scatter Plot of Observed Volume of Screenlines for Different Days	139

1. INTRODUCTION

1.1. Background

With the escalated challenges of congestion and constraints in building new roads, such as construction costs and right-of-way limitations, transportation agencies are increasingly implementing advanced operational strategies to maximize the performance of the existing infrastructure. Demand and access management, incident management, smart work zone applications, and advanced traveler information systems are examples of the types of strategies that are designed to get the most out of the existing physical capacity. Advanced technologies are needed to implement these strategies for the constant monitoring of traffic conditions, effective analysis of the traffic data in offline and online applications (for planning and operation), and active response to different traffic situations.

Intelligent Transportation Systems (ITS) have provided a solid platform for deploying the abovementioned strategies. The advances in ITS technologies and strategies have made collecting and archiving traffic data more efficient and affordable. This data can be used to closely monitor and analyze traffic conditions, in both real-time and offline applications, as well as to correspondingly plan, operate, and manage the facility.

Managed lanes (ML) are increasingly being considered as one of the most promising strategies to address transportation system problems. ML has evolved based on the notion of actively operating freeway facilities. A managed lane is a lane (lanes) within an existing freeway that can be dynamically managed to constantly meet preset criteria,

such as acceptable levels of service or minimum speeds. Advanced applications of managed lanes involve traffic management centers (TMCs) dynamically adjusting their operation parameters via controlling access to the ML, changing eligibility of vehicle occupancy, and varying the toll values to regulate the demand and keep the facility in optimal operational condition.

Effective planning and implementation of ML strategies require the utilization of advanced modeling methods to allow for a more accurate assessment of the impacts of changes in traffic flow conditions and the impact of operation strategies. Macroscopic, mesoscopic, and microscopic analyses and simulation have been used in assessing managed lane strategies. Mesoscopic simulation modeling has been proposed as a level of modeling detail between macroscopic and microscopic modeling since microscopic simulation is expensive to apply and calibrate, and macroscopic analysis is not capable of capturing the dynamics of traffic flow, particularly under congested conditions with breakdown and queue spillback effects. Dynamic Traffic Assignment (DTA), combined with mesoscopic simulation and in some cases, microscopic simulation, has been increasingly used to evaluate traffic management strategies. Compared to the traditional methods that normally utilize Static Traffic Assignment (STA) and simple analytical traffic flow equations, simulation-based DTA better captures the dynamics of system operations by modeling time variant system measures (including queuing and travel times), demand, advanced management strategies, and the associated responses of travelers.

Calibration is a substantial challenge in DTA utilization and significant research has addressed this issue during the last three decades. Calibration methods can be divided

into three categories: supply calibration, demand calibration, and joint supply-demand calibration. Supply calibration, also known as network calibration, identifies traffic flow and network parameters that result in replicating real-world measurements such as traffic volume, travel time and queue, given that the demand used in the model is highly reliable (Kunde, 2002). Demand calibration, also known as demand estimation or time-variant origin-destination (OD) matrix estimation, aims at estimating the OD matrices based on data from different sources by utilizing different methodologies, assuming that the network and traffic flow model is well calibrated (Zhou, 2004; Balakrishna, 2002).

A well calibrated network and a reliable OD matrix are not usually available beforehand. To account for the interrelation between network properties and travel demand, joint network-demand calibration procedure has been suggested that iterates between these two parts until convergence (Balakrishna, 2005; Gupta, 2005). Advances in computation power allow for efficiently estimating of all supply and demand parameters together. This approach, known as the simultaneous calibration approach, has been modeled either as a stochastic optimization problem or a state-space problem (Vaze, 2007; Balakrishna, 2007). The drawback of these fully automated methods is that local knowledge about the network cannot be easily incorporated in the process of calibration (particularly in stochastic optimization). In addition, allowing many parameters to change at the same time in the optimization complicates the calibration process significantly without assuring of successful implementation in real world.

The current research develops methodologies to support the development, calibration, assessment, and use of DTA in modeling managed lanes, and subsequently takes advantage of the presence of the detailed traffic data from ITS implementations.

1.2. Problem Statement

Successful implementation of ML depends on understanding and predicting the trip makers' choice to use these lanes. Agencies want to know what demand ML will attract, which can be translated into revenue, as well as congestion relief on the competing General Purpose lanes (GPL). They also need to know how the user will respond to changes in operation policies, such as changing the toll value or vehicle occupancy eligibility. Answers to these questions can be provided using advanced modeling techniques combined with effective demand estimation and validation methods.

In the following discussion and the rest of this dissertation, the reader should differentiate between two related terms for demand. The general term of “demand” refers to the number of trips that pass through the network as a whole, whereas the term “ML demand” refers to the number of trips that prefer to use ML. Initial demands can be obtained from a regional demand forecasting model but normally have to be significantly improved before using them as inputs to advanced modeling procedures. ML demands, on the other hand, are obtained through the modeling and calibration of route choice behavior.

The benefit of ML is most pronounced during rush hours, when trips get longer and less predictable due to congestion. Therefore, a good ML modeling framework should be time-dependent and sensitive to variation in demand and network properties. Although the superiority of DTA over traditional STA for applications such as ML has been discussed in the literature, a number of issues hinder the use of DTA in ML modeling, such as the data required for modeling and calibration being difficult to obtain. This is beginning to change, however, as these data are now becoming available in some

regions and will be more accessible and affordable in additional locations in the future. Nevertheless, the value of these data has been largely ignored, and there is a need to educate agencies on how to make the most out of these data in modeling and calibrating of ML applications. Therefore, the present study highlights the potential of data from a variety of sources for ML modeling and calibration.

Transportation modelers are still not sure of the benefits of DTA versus traditional STA methods when modeling managed lanes and how best to use DTA combined with detailed data and other estimation and calibration techniques for this purpose. There is a need to explore of different approaches to model route choice behavior and demand estimation and compare the results of the analysis. In addition, there is a need to develop and investigate methods to improve calibration accuracy and credibility, and to enhance demand estimation. Furthermore, there is a need to explore how different modeling steps can be combined to produce an effective modeling framework.

Below are tasks that need to be conducted to satisfy the above needs. These tasks were performed in this study.

Demand Estimation Procedure: The initial demand taken from regional demand forecasting models has to be improved before using it as inputs to ML modeling. A process is needed to predict more detailed origin-destination (OD) trip matrices that, when loaded onto the network, can better replicate real-world traffic conditions, while ensuring that their deviation from the initial OD trips is controlled and justifiable.

Managed Lane Modeling: There is a need to determine the degree that dynamic traffic assignment is better in modeling ML compared to traditional static assignment. Once this is established, there is a need to compare two of most common approaches of

modeling ML in the assignment step. In the first approach, the toll cost is converted to the equivalent travel time and is added to the generalized cost function of the link. In the second approach, prior to the traffic assignment, a willingness-to-pay curve is used to determine the percentage of travelers who are not willing to pay the toll. In the assignment step, this group of users is prohibited from using ML, and the rest of the assignment is governed by the user equilibrium principle, based on travel time on ML and GPL.

Calibration Process and Performance Measures: Proper calibration process and performance measures are necessary for ensuring the credibility of the result. Demand estimation, traffic flow model calibration, and assignment calibration aim at minimizing the discrepancy between the simulated network measures and corresponding measures based on real-world data. In some studies, this discrepancy has been merely limited to the difference between link volume and real-world traffic counts. Particularly in ML assignment, in which, the solution highly depends on the difference between travel times on ML and GPL, replicating congestion patterns should be considered as important as volume replication.

Solution Convergence: Convergence, stability, and proportionality of the assignment solution should also be checked. The problem with a calibrated but unstable network is that small perturbations in the network or demand attributes can cause an unreasonable response. Despite the emphasis on these criteria in the literature, they have not been properly addressed in most relevant studies.

Data Acquisition, Validation, and Processing: Recent advances in Intelligent Transportation System (ITS) technologies allow for collecting, archiving, and utilizing

valuable data to calibrate assignment models. Such data include: speed, count, occupancy and associated derived measures such as queue length and travel time by point traffic detectors, partial origin-destination and travel time data collected using Automatic Vehicle Identification (AVI), event data (such as incident and construction data), weather data, and recorded video by CCTV cameras. The integration of data from different sources may compensate for the limitations of each data source. As these data become more efficient to collect and archive, and more affordable to obtain, it becomes increasingly important to understand how valuable these data can be for improving the calibration quality.

1.3. Research Goal and Objectives

The goal of this dissertation is to develop and assess advanced strategies for managed lane modeling and calibration. The specific objectives are:

- 1) Develop a framework for using and assessing DTA modeling of managed lanes.
- 2) Develop procedures for calibrating and validating DTA modeling of managed lanes, utilizing detailed data from multiple sources.
- 3) Demonstrate the application of the developed procedures to real-world ML situations.

1.4. Dissertation Organization

The remainder of this dissertation is structured as follows: Chapter 2 presents a review and assessment of past research related to the objectives of this study. First, a review is presented of existing ML modeling frameworks. Second, DTA is introduced as

a potential method for ML assignment. Next, current practices and research in supply and demand calibration of assignment models are reviewed. Lastly, existing literature on model convergence is reviewed, illustrating that achieving a stable and equilibrated solution is important for the ML modeling.

Chapter 3 presents the methodology and tasks that have been accomplished throughout this study. This chapter explains the task sequence and the links between different tasks.

Chapter 4 describes the required procedures to prepare the demand and network from a regional demand model for DTA applications. There was a unique opportunity in this research in terms of accessing the ITS data-rich environment. Removing non-representative day and time intervals, removing detector erroneous data, and checking spatial and temporal consistency of data are crucial tasks for preprocessing and validating this data.

Chapter 5 describes the procedure for network (supply) calibration. The goal is to estimate capacity and traffic flow model parameters for network links. This process starts with replicating isolated bottlenecks and is extended gradually to cover a larger network.

Chapter 6 includes the framework for demand estimation consisting of sub-elements that can run sequentially in an ascending level of detail and complexity. In this chapter, proper performance measures are set to assure a reliable, reasonable estimation of demands.

Chapter 7 is dedicated to evaluating two approaches for utilizing DTA to assess managed lane modeling. The first approach is to incorporate the toll cost as equivalent time in the link generalized cost function. The second approach is to divide people to

toll-payers and non-toll-payers based on “willingness-to-pay curve” prior to the assignment. These approaches have been compared from different points of view. Static and dynamic assignments are also compared in terms of replicating real-world travelers’ behaviors in choosing ML, as well as model convergence and the stability of the assignment solution.

Chapter 8 summarizes the findings of this research on demand and supply calibration, and the assignment module in the context of managed lane modeling.

2. LITERATURE REVIEW

This chapter presents a review and assessment of past research related to the objectives and tasks of this study. First, a review is presented of existing ML modeling frameworks that were found to vary in their levels of details and complexity. Secondly, a brief introduction to DTA concept is provided, demonstrating the benefits and necessity of DTA utilizations in applications like ML assignment. Next, current practices and research in supply and demand calibration, separately or jointly, are reviewed, leading to the selection of an iterative-joint approach to supply and demand calibration for use in this dissertation. Lastly, existing literature on convergence are reviewed, illustrating that achieving a stable and equilibrated solution is important to ML modeling that requires assessing different strategies relative to one another. Without assuring a stable and well-converged network, it is not possible to differentiate between differences in performances that are due to changes in inputs and policies and those that are due to model noise and randomness of the non-converged models.

2.1. Managed Lane Modeling Frameworks

A variety of modeling approaches have been proposed to assess managed lane implementations. These approaches range from high-level sketch planning tools to micro-level modeling of individuals' behaviors and traffic operations.

The Federal Highway Administration (FHWA) developed an open source sketch planning tool (POET-ML) to perform a quick evaluation of ML functionality and pricing policies. The input into this spreadsheet includes eligibility policies such as occupancy restrictions; physical characteristic such as lengths and numbers of the lanes, median

types, and buffer types; and demand information such as the peak hour volumes on ML and GPL facilities. The user can change the current policy according to the results produced by the tool, and can also review the potential impacts on travel demand, revenue, mobility, and the environment (FHWA, 2008a).

TRUCE 3.0 and TRUCE-ST are similar tools developed by FHWA that allow the user to quantify the impacts of congestion pricing on urban highways at the State level. The input includes aggregated traffic data from urban mobility reports (Schrank, and Lomax, 2007) and socioeconomic data from census for the desired study area. The tools allow for the evaluation of the effect of different congestion pricing policies on traffic condition, air quality, and revenue (FHWA, 2008b).

FITSEVAL is another sketch planning tool developed for the Florida Department of Transportation (FDOT) by Florida International University in Miami, Florida, to evaluate and assess ITS alternatives within the Florida Standard Urban Transportation Model Structure (FSUTMS) framework. This tool evaluates the effects of ITS applications including ML, on network performance measures such as Vehicle Mile Traveled (VMT) and Vehicle Hour Traveled (VHT), average speed, and fuel consumption (Xiao et al., 2010).

When utilizing the four-step demand forecasting framework to model ML, the most straightforward approach is to add a toll term to the generalized path cost in the assignment module and assign a dollar value to travel time in the generalized cost function. Recently, travel time reliability was also added to the generalized cost function of the link. More advanced models apply a binary route choice (toll vs. non-toll routes), either within the assignment or externally, tying this binary choice to the assignment in

an iterative manner. Recent applications have modeled the travelers' behaviors in choosing ML by utilizing probabilistic approaches, such as using a logit model based on a derived utility function or a willingness-to-pay distribution based on traveler surveys.

An essential component of the managed lane choice, whether implemented in the generalized cost function of the assignment process or as a separate logit model, is the Value of Time (VOT). VOT is a measure of a driver's willingness-to-pay for travel time savings. VOT is a means of capturing dissimilarities between different classes of drivers in route choice; more specifically, either in mode choice, route choice, or within assignment. These dissimilarities are caused by several socioeconomic and trip factors. Chiu (2012) compared modeling VOT dissimilarity in discrete choice model versus multi-class stratified assignment. In the discrete choice model, every traveler makes decisions of choice based on a generalized utility function (GPL vs. ML), while in the multi-class stratified assignment, predefined classes of travelers (stratified based on VOT distributions) are assigned separately. The first approach is time-efficient and easier to implement, but difficult to converge. The second approach is more time-consuming, but produces a more stable solution; therefore, it is more appropriate for scenario comparison applications. The toll choice procedure in the Southeast Regional Planning Model (SERPM) is conducted utilizing the standard multimodal logit model, and is conducted separately for each trip purpose and each vehicle occupancy category. (FDOT, 2013)

More comprehensive models consider the toll and VOT (and potentially the value of reliability or VOR) in the utility function of the mode split and the impedance function in trip distribution. In these cases, linking different steps (assignment, mode split, and/or distribution) is essential to ensure consistency between their outputs. After partitioned toll

and non-toll trips are calculated and loaded into the network in the assignment module, the travel time is skimmed and fed back into the mode choice and trip distribution steps. Mutual consistency should exist between different levels. For example, global convergence problems were reported for cases where the toll cost is modeled in the assignment generalized cost function and in the mode choice, but not in the impedance function used in the trip distribution (NCHRP, 2012).

Boyce et al. (2008) mentioned that the travel time input to trip distribution and mode choice should be equal to the travel time (cost) obtained from the equilibrium assignment in the next step. The author performed several computational experiments of how to incorporate the feedback into demand forecasting models. It was found that the direct (naïve) feedback is not efficient, and a type of averaging is needed. In comparing different alternatives of what to average and how to average, it was recommended to average the trip matrices with fixed weights (e.g., weights do not change by iterations). The converged solution will produce a matrix, that when loaded onto the network by the assignment module, generates route travel times that if fed back to trip distribution and mode choice step, would reproduce the same matrix. The same feedback procedure is applied in the enhanced demand forecasting model in Florida to overcome inconsistency issues between trip distribution/mode choice and assignment (FDOT, 2013).

There has been a recent interest in utilizing Activity-Based Models (ABM) in managed lane studies. It has been argued that traditional trip-based models are unable to respond to pricing policies in trip generation, departure time, and occupancy choices. In both approaches (trip-based and activity-based), there is a growing trend of combining

mode, occupancy, and binary route choice in a multi-level nested logit model structure when modeling ML (Vovsha et al., 2013)

Recently, a survey was conducted as a partial effort to incorporate toll modeling into the existing Phoenix metropolitan area demand model by URS (URS, 2011). This survey aimed to identify the best practices of toll modeling. Seventeen agencies responded to this survey. Sixteen of these agencies currently use the four-step demand forecasting model, with six agencies planning to replace the conventional trip-based model with ABM. Nine metropolitan planning organizations (MPOs) incorporate the toll cost in the impedance function in the trip distribution step. The Nested Logit model is the most commonly used mode choice model. Nine agencies partition the trip table between toll and toll-free users, either in their mode choice or assignment model.

A majority of agencies responding to the abovementioned survey used a feedback loop, from trip assignment to trip distribution, or to mode choice. In almost all cases, the assignment method is static user equilibrium. Seven MPOs consider both travel time and toll cost to calculate the shortest paths. The route choice model, both as a sub-element of the mode choice hierarchy and in the assignment step, has the advantage of sensitivity to socioeconomic characteristics. A calibrated logit-type model or willingness-to-pay distribution can be used in the assignment module to define the route choice behavior. The final prototype proposed by URS was an advanced highway assignment with a customized route choice that feeds back to trip distribution and mode choice. The utility function takes into account the income levels and bias factors. A bias coefficient accounts for unknown factors that affect single occupant vehicle (SOV) decisions, such as perceived improved trip reliability, safety, and comfort.

A key step in ML design and modeling and predicting the associated demand and revenue is to estimate VOT (and potentially VOR) and associated factors that affect a traveler's decision to choose or avoid paying a toll. VOT is generally referred to as the monetary toll value divided by the saved time, or equivalent "perceived" benefit for using ML. The heterogeneity of travelers is a crucial property to be captured. The necessary level of model detail requested by the National Environmental Policy Act (NEPA) for traffic and revenue (T&R) analysis includes: four to five major travel purposes, three to four income groups, and three to four time-of-day periods. Vovsha et al. (2013) recommended considering the length of trips and congestion levels in VOT estimation. It was found that drivers perceive every minute in congestion as 1.5 to 2 minutes of free-flow driving.

Recent findings recommend including travel time reliability as a decision factor in the assignment process, and subsequently, VOR was introduced in the generalized cost/utility function. Two general approaches are introduced in measuring travel time reliability. The first approach relates reliability to variability, meaning the higher variability in travel time (measured as trip travel time variance or similar concepts) is equivalent to a less reliable trip. The second approach measures reliability as a portion of success or failure against pre-established thresholds, such as proportion of trips with a delay less than a predefined threshold (Cambridge Systematics, Inc., 2012).

In order to obtain travel time reliability from stated and revealed preference surveys, the Resource Systems Group (2012) associated travel time reliability with travel time distribution entropy. It is assumed that travelers will pay to reduce the entropy.

The entropy is calculated as a function of the mean and standard deviation of the travel time distribution. The value of reliability is in dollar per unit of entropy.

Minnesota was the first state to implement a fully dynamic pricing algorithm that updates the toll based on High Occupancy Toll (HOT) lane density and density variability every three minutes, with a goal to keep the level of service at C (Janson and Levinson, 2013). By implementing different toll policies and analyzing the flow on the ML, a counter-intuitive positive correlation between pricing and ML demand was observed. The authors believed this contrary behavior is because drivers perceive the toll value as an indication of GPL congestion level. Similar results were observed in Burris et al. (2012). The authors performed data analysis on two HOT lane facilities in Minnesota and California, which revealed that in Minnesota, SOVs pay up to \$116/hour, and in California up to \$54/hour to use HOT lanes during the afternoon peak, and slightly less for the morning peak. The authors interpreted these high values are not only paid toward time saving, but also for improvements beyond time saving, such as trip reliability. Alvarez's (2012) research at Florida International University showed that based on historical ITS data, people occasionally chose to pay toll during the AM peak, while the parallel GPL had a lower travel time. The Resource Systems Group (2012) showed that the saved time in ML is overestimated by travelers, by comparing joint stated and revealed preference surveys with historical data.

Much lower values of VOT are used in practice and are recommended for modeling as default values (NCHRP, 2012; Vovsha et al., 2013). An Investment Grade Traffic and Revenue Study (WilburSmith, 2011) reports a range of \$6/hour to \$18/hour of VOT, with an average of \$14.31/hour for the US 36 Corridor in Colorado. Past studies

have shown that two groups of factors affect SOV decisions to use ML: 1) Trip-related factors such as trip length and purpose, trip time of day, travel time savings, improved trip reliability, safety, and comfort; and 2) Socioeconomic factors such as income level, age, gender, and household composition. It was found that the income level and trip purpose are the most influential factors (Burriss et al., 2012).

For calibration of the ML model within the Florida SouthEast Region Planning Model (SERPM) framework a value of time of \$1 equal to 5.1 minutes (\$11.75/hour) for VOT, and a range of \$0.00 to \$2.99 for VOR are suggested. These values are based on stated and revealed preference surveys in fall 2011 (Resource Systems group, 2012). Calibrating models based on stated and revealed preference surveys for the Florida Turnpike's tolling framework has resulted in a VOT ranging from \$3/hour to \$13.50/hour, based on trip purpose and income level (Dehghani et al., 2003). Nava et al. (2013) selected a VOT of \$15.50/hour for SOV and HOV users and a VOT of \$46.50/hour for commercial trucks. In their methodology, the toll value update mechanism is internally implemented within a dynamic user equilibrium framework, which implies mutual consistency and convergence between toll value and route choice.

Choosing ML versus GPL is a learning process for commuters. Studies show that the learning process that leads to a high correlation between saved time and ML selection takes about 60 days. In other words, it takes 60 days of adjustment prior to choosing ML over GPL, based on the saved travel time (Alvarez, 2012).

Sometimes, constants are also included in the utility functions to account for unobserved factors that lead travelers toward ML or away from it. These parameters are hard to measure and are estimated through model calibration and fine-tuning tasks. In the

SERPM model calibration, bias against the HOT lane choice is inserted in the utility function for off-peak periods. For peak periods, a bias toward HOT lane is included to replicate the observed volume on the HOT lane. In addition, saved travel time is exponentially increased with congestion level to reflect how travelers “perceive” the benefit of using HOT versus GPL when the road is heavily congested. This effect was revealed in travel surveys and stems from better safety, comfort, and reliability when using HOT (FDOT, 2013).

It should be noted that the revealed VOT for ML might be different from the VOT for toll facilities when the entire facility is tolled (e.g. Florida Turnpike). This is because with ML, drivers can decide at the last moment which route to take based on dynamically changing traffic conditions and tolls. Moreover, with ML, usually a small portion of ML capacity can be purchased by SOVs, therefore, only SOVs with relatively high VOT will divert compared to toll facility users.

2.2. Dynamic Traffic Assignment

The impacts of advanced strategies such as ML are particularly significant when the facility is operating near its capacity. Applying these strategies is time-dependent and highly sensitive to small changes in traffic and/or demand. Therefore, these applications require more advanced and detailed modeling frameworks, compared to the approaches used in traditional demand forecasting.

The use of simulation-based DTA was proposed as an alternative to STA to provide more realistic and detailed analyses of ML. Simulation-based DTA tools utilize mesoscopic or microscopic simulation to assess traffic performance after each assignment

iteration. Mesoscopic models generate and track individual vehicles, as is the case in microscopic simulation, however, the interaction between them is modeled through a macroscopic traffic flow model (TFM), rather than a microscopic traffic modeling. Microscopic simulations are useful tools for traffic analysis. However, they are extremely demanding in the data and time needed for the correct modeling and sound calibration of traffic flow model. Such models are not appropriate for regional networks. Macroscopic simulations, on the other hand, are too aggregated for operational analysis purposes, and many are unable to capture vital features of congested networks like bottlenecks.

DTA is a modeling approach that captures the dynamic interaction between demand and network, and advanced strategies and associated parameters. It models the period demand over short-time intervals, with a traffic assignment in each interval, which is affected by the network condition resulting from the previous interval assignment. This means that for each OD pair, vehicles that depart in different time intervals can use different paths and may experience different travel times. The core engine that assigns the demand to eligible routes in most static and dynamic assignment tools stems from the user equilibrium (UE) concept. Equilibrium means that for each OD pair, the experienced travel time on different routes are the same, and no traveler can improve his/her travel time by switching the routes. In DTA, dynamic user equilibrium is to be achieved for every departure time interval.

To better understand the difference between STA and DTA, it is necessary to first understand the main components of traffic assignment procedures that run sequentially and iteratively seeking a convergence. These three main components are:

- Shortest path identification (also referred to as tree-building): This includes the identification of a set of attractive paths (routes) between each OD pair. In DTA, this component is time-dependent and includes updating the set of attractive paths given the estimated travel times of the paths during the previous assignment process.
- Assignment of the trip demands to the identified attractive paths: This component results in the estimation of link flows by assigning the demands to the competing attractive paths. In DTA, the proportions of demands assigned to each path are calculated for each assignment time period. In general, a time period of 15-30 minutes is most widely used.
- Network loading: This component refers to the representation of the movement of vehicles on the network as they travel from origins to destinations. Network loading allows the estimation of performance measures for use in the assignment, such as route travel time between origins and destinations. In DTA models, network loading procedures can be classified as analytical procedures or simulation procedures. Due to the complexity of traffic operations, particularly with the presence of congestion and traffic control, simulation-based procedures are the most widely used types of procedures at the present time (Hadi et al., 2012).

The discussion above indicates that unlike STA, which defines the shortest paths and allocates all of the traffic to these paths at once for the whole peak period, DTA conducts the traffic assignment and reaches equilibrium for each time interval far shorter than the model period. This is preferred in two aspects, as follows: 1) DTA can model

time variant demands, time variant operational strategies (such as those applied in ML), associated travelers' responses, dynamic variations in network performance, and dynamic events such as lane blockage incidents; and 2) Simulation-based DTA can model queue building and dissipation and queue spillback due to exceeding link capacity or downstream link queuing capacity, as it occurs in the real world. Therefore, DTA provides a more realistic representation of travelers' behaviors and traffic conditions, and provides a better approach for assigning traffic and estimating travel cost and time, resulting in better demand and performance measure forecasting.

Despite the potential benefits of utilizing DTA, there are some concerns and issues hindering its use. The most common concerns identified by modelers and planners include 1) the excessive data and time needed to model and calibrate DTA networks, 2) the required time and cost for training, and 3) the time required to integrate DTA with other transportation analysis tools such as demand forecasting models, multi-resolution modeling, and ABM modeling. In particular, integration of DTA with activity-based or choice models is difficult to converge.

In April 2009, the Transportation Research Board (TRB) Network Modeling Committee conducted a DTA user survey through the Federal Highway Administration (FHWA) Travel Model Improvement Program (TMIP) mail list, which shows that more than 70% of the 85 respondents plan to apply DTA tools within two years (Tung and Chiu, 2011). On the other hand, the respondents also clearly identified the following top five technical and institutional barriers:

- DTA requires more data than current availability or accessibility (47%)
- Setting up a DTA model takes too many resources (44%)

- Cost/benefit is unclear (45%)
- DTA tools take too long to run (35%)
- Modeling approaches are unclear (35%)

Another survey was conducted in 2010 by the Florida modeling community, related to their views of DTA applications and limitations. Forty-seven responses were received from private sectors, metropolitan planning organizations (MPOs), and state agencies. Thirty-six percent of responders believed that there is a lack of data for DTA applications at this stage of development; 24% mentioned lack of experience as an obstacle for DTA implementation, 22% were concerned about calibration and validation requirements, and 21% named computational time as a DTA drawback compared to the traditional regional models. The need for training, complexity of the process, and the cost of software were also confirmed as issues when considering implementation of DTA (Hadi et al., 2012).

Convergence of DTA models should also be an important area of consideration by modelers. In static user equilibrium, the convergence of the solution is theoretically provable. However, in simulation-based DTA tools, the convergence is not theoretically guaranteed. Therefore, arbitrary performance measures are introduced as convergence criteria, with no agreed-on acceptance levels.

2.3. Supply/Network Calibration

Supply calibration includes the estimation of parameters associated with traffic operations in the network. These parameters vary depending on the type of the model (macroscopic, microscopic, or mesoscopic) and the specific tool under consideration. The

parameter used in mesoscopic simulation tools generally include segment capacities, free-flow speed, queuing density and/or jam density, and/or other parameters used in the macroscopic traffic flow model used to move the vehicles onto highway segments. The performance of the system with the selected parameters is evaluated by comparing the model results to real-world measures of traffic flow, such as queue formation and spillback, density, and travel time on each link.

2.3.1. Mesoscopic Simulation Supply Calibration

Kunde (2002) calibrated the network supply of the DynaMIT model through a sequential process at increasing levels of aggregation. The process starts at the level of separated bottlenecks where capacity is estimated by various methods based on field data. The network is gradually extended to connect the bottlenecks, and then model the whole corridor. The parameters from the previous steps are fine-tuned, and the supply-demand calibration runs iteratively until a desirable convergence is achieved. The most disaggregated level is the individual segment level, at which the speed-density relationship and capacity are calibrated. At this stage the interactions between adjacent segments is ignored. Due to the lack of data and large number of variables, network segments were first grouped into 11 representative clusters. All segments in a cluster were set to have the same TFM parameter values. The next stage is to perform calibration at the sub-network level where the origin-destination flows can be reasonably estimated solely from the sensor counts, because the probability of a second alternative route choice between each origin and destination is zero or negligible. This way, the impacts of errors in demand estimation on supply calibration are deleted. The last step is the network-wide

calibration, which takes into account all of the interactions between various segments and any errors due to demand estimation. Stochastic optimization is used to calibrate the supply at the whole network level.

To estimate the macroscopic TFM model parameters in Dynasmart-P with the modified Greenshields model as utilized TFM, Mahmassani et al. (2004) rewrote the model formula in the natural logarithmic form, whereby the relation between speed and density becomes linear. The authors estimated the parameters by performing multiple runs of regression analysis. In each run, they set one of the parameters as fixed and systematically changed the other parameters within a reasonable range to determine the optimum combination that replicates detector data.

Wang et al. (2009) applied Kalman filtering to continuously estimate the state of the traffic based on real-time data. Capacity and TFM parameters were calculated within a stochastic nonlinear macroscopic TFM framework by an adaptive estimator. This method does not require an initial estimation of the parameters; it automatically adapts to changes in the model due to changes in external conditions and can recognize interruptions due to incidents. The drawback of this method is that the output cannot be related to the theoretical aspects of traffic flow.

The Highway Capacity Manual (TRB, 2000; TRB, 2010) is used as the authoritative source of defining and estimating capacity in the United States. A procedure is presented in the HCM that allows estimating freeway capacities based on free-flow speed. The procedure allows adjusting the capacity estimates to account for deviations from default conditions, considering a limited number of factors. However, many other parameters affecting capacity are not considered in the adjustment. Thus, the HCM

encourages measuring capacity in the field to consider the differences in geometry and driving characteristics between different regions and facilities.

The remaining subsections of Section 2.3 discuss in more detail the specific aspects of the calibration process, including bottleneck identification, free-flow estimation, capacity, and TFM parameter estimation.

2.3.2. Bottleneck Identification

In a congested network with recurrent bottlenecks, the most crucial part of network calibration is to replicate bottlenecks as they happen in the real world, in time and space, and correctly estimate the capacity and impacts of the bottleneck.

A bottleneck is defined as a point upstream of which a queue is formed, with the traffic flowing at free-flow speed at downstream locations (Bertini et al, 2008). Bottlenecks can be active or hidden. A hidden bottleneck is a potential one that is a result of geometric or demand features but cannot be observed because the approaching traffic demand is metered by another upstream bottleneck. An active bottleneck is the only location where capacity can be measured based on field data. Chen et al. (2004) identified bottlenecks based on the speed differences between adjacent detectors, where the speed at the upstream detector is below a particular threshold (e.g., 40 mph), and the speed drop is above a particular threshold (e.g., 20 mph). The required parameters, including the maximum speed threshold, minimum speed difference between adjacent detectors, and data aggregation levels were recommended to be site-specific. Zhang and Levinson (2004) identified bottlenecks based on the occupancy differences between adjacent detectors. Hall and Agyemang-Duah (1991) used the occupancy-to-flow ratio as a

bottleneck identification criterion. Bertini and Myton (2005) used cumulative vehicle counts and cumulative occupancy graphs to identify bottleneck activations without the need to set speed or occupancy thresholds.

2.3.3. Free-Flow Speed

Free-flow speed (FFS) is a crucial parameter in the HCM capacity estimation procedure for uninterrupted facilities. The HCM provides a free-flow speed estimation procedure that incorporates reduction factors to account for deviations from base conditions. Reductions in free-flow speed will implicitly drop the capacity, according to the HCM procedure.

Equations 2-1 and 2-2 show the relationship between the basic and adjusted free-flow speed to account for the deviations from basic conditions in the HCM 2000 and HCM 2010 respectively.

$$FFS = BFFS - f_{LW} - f_{LC} - f_N - f_{ID} \quad (2-1)$$

BFFS= base free-flow speed (75 mph for rural freeways and 70 mph for urban freeways),

f_{LW} = adjustment factor for lane width (mph),

f_{LC} = adjustment factor for right shoulder lateral clearance (mph),

f_N = adjustment factor for number of lanes (mph), and

f_{ID} = adjustment factor for interchange density (mph), and

$$FFS = 75.4 - f_{LW} - f_{LC} - 3.22TRD^{0.84} \quad (2-2)$$

TRD = total ramp density (ramp/mi).

The HCM encourages users to measure FFS in the field as the average of all vehicle speeds when the volume is less than 1000 pc/lh/hr. Chao et al. (2005) used the average of speeds when occupancy is below 10 percent. Dervisoglu et al. (2009) estimated FFS by fitting a straight line to the uncongested part of the fundamental diagram.

2.3.4. Capacity Definition and Estimation

The HCM defines freeway capacity as the maximum sustained 15-minute flow rate that can be accommodated by a uniform freeway segment under prevailing conditions. As mentioned earlier, the HCM recommends values of capacity based on free-flow speed, and provides a few adjusting factors to account for deviations from prevailing conditions. However, there is evidence that these adjustments are not enough to reflect the significant differences between locations due to geometry, demand, and driving characteristics (Washburn et al., 2010). Given a determined FFS and weather condition, heavy vehicle and driver population are the only factors used to adjust the capacity. The heavy vehicle percentage can be obtained by detectors that classify vehicles, or by manually counting vehicle classes. However, the driver population, which is the percentage of non-commuters that are not familiar with the analyzed highway, is very difficult to estimate.

To account for site specifications, direct measurements of capacity were recommended. In absence of a recommended method by the HCM, researchers proposed a number of approaches for these measurements. Dervisoglu et al. (2009) estimated capacity as the maximum observed 5-minute flow rate over several days. Chao et al. (2005) estimated the capacity as the maximum hourly flow observed during a 30-day period. Jia et al. (2010) estimated capacity as the average of the top one percentile

of a 15-minute flow rate over several days, which turned out to be similar to values estimated by the HCM.

Van Arem and Van der Vlist (1992) estimated capacity by determining the maximum occupancy in the uncongested part of the fundamental traffic flow diagram and the associated volume. Bassan and Polus (2010) approximated the capacity by fitting data into parabolic speed-flow and flow-occupancy models. Similarly, Wang et al. (2009) used the apex of a flow-density curve as capacity. Rakha and Arafeh (2010) performed an automated fitting procedure of a quadratic speed-flow function to loop detector data. This function combines the microscopic Pipes car-following model and the single regime Greenshields model. The automated model calibration yields an estimated number of key parameters, including capacity.

Researchers have also argued that capacity is not constant, even under identical external conditions (Elefteriadou et al., 1995; Minderhoud et al., 1997). These researchers recommended a paradigm shift in capacity calculation, from a deterministic value to a stochastic value, and proposed statistical methods to measure capacity. In most of these studies, capacity is tied to the notion of traffic breakdown. The most common proposed values as capacity representatives are queue discharge flow and the maximum flow before breakdown. The queue discharge rate is defined as the long-run average of flow over the breakdown period. Pre-breakdown flow was measured using different time intervals before breakdown, such as 5 minutes and 15 minutes (Elefteriadou and Lertworawanich, 2003; Hall, and Agyemang-Duah, 1991).

Based on a lane-by-lane analysis of breakdown, Dehman (2012) pointed out that in some cases, the flow increased after the breakdown and explained that this mainly

happened because of lane changing between underutilized and fully utilized lanes. Brilon et al. (2005) found that a freeway operates at the highest expected efficiency only if it is loaded to 90% of the conventionally estimated capacity.

There are no guidelines on whether to use pre-breakdown, queue discharge, or a weighted combination of both as values representing capacity (Zhang and Levinson, 2004). It has been reported, however, that queue discharge is lower than the pre-breakdown flow by 2 to 26 percent in different studies, mostly due to a change of driving behavior to stop and go status (Yeon et al., 2007; Hall, and Agyemang-Duah, 1991). The HCM 2010 also recognizes this phenomenon; however, it does not consider it in its procedures and does not recommend any specific percentage of capacity reduction after traffic breakdown.

A freeway facility HCM computational engine was developed to implement the HCM 2010 Chapter 10 procedure, so as to estimate freeway capacity when queue exists. In this engine, called FREEVAL, oversaturated conditions are followed by a user-defined drop in capacity, reflecting the queue discharge rate during these conditions. The National Cooperative Highway Research Program (NCHRP) project 3-96 also aimed to develop methods for the performance assessment and capacity analysis of managed lanes compatible with HCM procedures. The result of this project is the development of additional features in FREEVAL, resulting in the FREEVAL-ML package that allows modeling of the GPL and the parallel ML (Wang et al., 2012).

In more recent studies, to account for the probabilistic nature of capacity, some researchers recommend calculating it as a percentage of the breakdown probability distribution. The most common utilized probability functions are the normal and

Weibull distributions (Hall, and Agyemang-Duah, 1991; Elefteriadou and Lertworawanich, 2003; Brilon et al., 2005). Minderhoud et al. (1997), which state that given a true distribution of capacity, one can obtain the capacity value by choosing the average, median, or 90th percentile of the distribution. This choice so far has been arbitrary and supported by the results from testing the local data goodness-of-fit. There is no consensus on which point of the breakdown distribution should be used to estimate capacity. Lin (2009) used bi-level linear programming to exclusively calibrate capacity in a DTA model. The upper level problem minimizes the deviation of simulated and observed occupancy data, and the lower level runs a simulation-based cell transition assignment.

In summary, the HCM is regarded as the most reliable source for estimating capacity for different facility types. However, the HCM procedures allow for the use of a number of factors to reflect local conditions. In some cases, however, this adjustment may not be sufficient, and direct measurement of capacity is needed. A variety of surrogate measures have been proposed for capacity measurements. It is worth mentioning that in some studies, the measured capacities reported as being lower than those estimated by the HCM (Washburn et al., 2010).

2.3.5. Traffic Flow Model Parameter Estimation

DTA tools use mesoscopic simulation models to generate and track individual vehicles, but move vehicles according to macroscopic relationships that are subject to link capacity and link storage limits. Depending on the specific model under consideration, the utilized macroscopic relationships could include the Bureau of Public

Roads (BPR) relationship, the modified Greenshields model, the Van Aerde model, or the Akcelik model.

BPR is the most common model in traditional STA-based demand forecasting applications. It has also been used in the Cube Avenue DTA tool (Citilabs, 2013). Different values have been suggested by practitioners to calibrate the BPR curve parameters to better replicate observed performance measures such as speed, volumes, total Vehicle Miles Traveled (VMT), and Vehicles Hours Traveled (VHT). In some applications, the parameters are set based on facility type and design speed. In more advanced applications, a volume/capacity (v/c) threshold is selected to divide the BPR curve into two different regions with different coefficients to reflect the difference in traffic dynamics between these two regions. The v/c values of 1, 2, and 4 have been used as thresholds in different studies (Spiess, 1990a; Singh, 1995; Dowling, 1997 and Hansen, 2005).

Saberi (2010) compared the results from the HCM empirical speed-density curves, BPR formula, and Davidson formula (Davidson, 1966 and 1978) and its descendent, the Akcelik formula (2003), and assessed their abilities to replicate the observed speed-density curves. The author recommended the use of the BPR curve for $v/c < 1$ and Akcelik formula for $v/c > 1$, since this formula accounts for the presence of queue. The author incorporated the probability distribution of capacity into the speed-density relationship to account for the stochastic nature of capacity.

Huntsinger and Roupail (2011) improved the accuracy of the BPR, Conical, Akcelik and HCM traffic flow models by replacing the volume with the estimated demand in these TFMs. The demand is calculated as the summation of volume at

capacity and queue at the bottleneck location. The authors optimized the parameters of the abovementioned TFMs to fit the demand/capacity versus travel time observations.

Dervisoglu et al. (2009) presented an automated empirical calibration approach of TFM parameters for a cell transmission model. The TFM is formulated as a triangular relation between flow and density. Capacity is estimated as the maximum 5-minute flow rate over several days. This value of flow on the flow-density curve is then projected horizontally to meet the free-flow speed line (a line from the origin of the diagram with a slope equal to free-flow speed) to establish the tip of a triangular fundamental diagram. This point corresponds to the critical density, above which the flow is considered to be congested.

Van Aerde and Rakha (1995) performed an automated fitting of a quadratic speed-flow function. This function combines the microscopic Pipes car-following model (applied in CORSIM) and a macroscopic single regime model (the Greenshields model). Speed and volume (and density if available) measurements from detector data were used to calibrate four parameters that define the relation between speed and density.

Chiu et al. (2010) introduced a vehicle-based mesoscopic model called the Anisotropic Mesoscopic Model. Instead of using the conventional TFMs that assume the same speed for all vehicles on a link at a given time step, vehicles on a link can travel at different speeds. In this model, the speed is affected by the presence of leading vehicles within a neighborhood, called the speed influence region (SIR).

Loudon (2007) pointed out that the traffic characteristic is quite different in ML, compared to GPL. In particular, the observed ML speeds were found to be lower than the original estimations, depending on the degree of separation between ML and GPL. This

is due to the interaction between ML and its adjacent, more congested GPL lanes. This effect is referred to as “side friction,” the degree of which depends on the separation type. The most significant effect was observed with marker painting buffers, and the least significant was observed with concrete barriers.

It is not feasible to estimate the capacity for every link when estimating the capacity in the field; first, because capacity can only be observed at critical link locations. This requires grouping road segments, which significantly reduces the size of the parameter estimation. Clustering can simply be based on geographical features such as number of lanes, horizontal/ vertical curve, and closeness to ramps (Balakrishna, 2007; Kunde, 2002), or through machine learning approaches such as the k-means algorithm.

2.4. Demand Estimation

Time-dependent origin-destination matrices are essential input to trip-based DTA models. Because of the very high cost of travel surveys, possible errors with these surveys, such as misreporting the trips and the need for a fine-grained demand matrix covering short-time intervals, methods must be developed to estimate reliable fine grained trip origin-destination (OD) matrices based on initial seed OD matrices obtained from demand forecasting models. Although seed OD matrices are very important in the estimation process, other sources of data, such as traffic counts and possibly partial OD matrices measured using Automatic Vehicle Location (AVL) or Automatic Vehicle Identification (AVI) data, are needed to improve the accuracy of the estimated matrices. The OD estimation methods can be categorized as assignment-based and non-assignment-based. Non-assignment-based methods apply traffic conservation

relations between entrance, exit and mainline volumes. These methods are mostly limited to road facilities without signals and without any queues. Other sources of information, such as AVI, are also difficult to incorporate into the models.

In general, the problem of OD estimation is underspecified, which means that the number of equations based on traffic counts on links are far less than the number of unknowns (OD table cells). Thus, different combinations of OD pairs can produce the same set of link volumes if loaded onto the network. To circumvent the problem of under-determinacy, researchers may aggregate ODs over longer time intervals, compare them to surveillance data time intervals, or alternatively, disaggregate the surveillance data into shorter time intervals. (Tavana, 2001; Gupta, 2005)

Assignment-based models utilize traffic assignment to map OD matrices to link volumes, allowing for the minimization of the deviation between model outputs and observed or estimated measures (such as initial OD matrices and measured traffic volumes) (Chi, 2010). Different sources of data are easy to incorporate into assignment-based models. In addition, if dynamic assignment is used, queues and signal delays are modeled by the DTA simulator. Thus, they are more appropriate to use than non-assignment-based estimation. However, the quality of the results of assignment-based models depends on the availability of high quality initial OD matrices (Lin, 2006). A main interest of this study was the current work being performed on the assignment-based OD estimation processes and the factors affecting this estimation.

Traditionally, assignment-based OD estimation is modeled as an iterative bi-level optimization, where the upper level minimizes the deviation between observed and simulated quantities, and the DTA simulator at the lower level produces a link-flow

proportion matrix as a result of loading the OD over network links. The simplest structure for this approach is depicted in Equation 2-3.

$$D = \arg \min_{\hat{c}} \sum_l \sum_t f(c_{(l,t)} - \hat{c}_{(l,t)})^2 \quad (2-3)$$

s.t.

$$\hat{c} = \hat{p} * D$$

and other sets of constraints, as discussed below.

In Equation 2-3, D is demand, and c and \hat{c} are observed and estimated traffic counts. The link with the detector measurement is l , t is the time interval with traffic data, and \hat{p} is link-flow proportion matrix that indicates which portion of each OD pair travels on a certain link. This matrix is usually obtained as a result of DTA modeling. The objective function is not limited to minimizing the deviation between simulated and observed counts. It can be extended to consider the deviation between simulated and observed speed, density, queue length, or the distance between an initial set of demands (seed OD matrices) and the estimated demands. Constraints also include, but are not limited to, non-negativity constraints, initial values, link capacities, cordon line counts, fixed OD flows, and/or production/attraction counts. Even route choice probabilities can be used as constraints if these parameters are to be fixed.

Tavana (2001) modeled the upper level of the OD estimation problem as a generalized least square (GLS) optimization to minimize the discrepancy between the estimated and measured link volumes. In favor of GLS, Brandiss (2001) pointed out that GLS formulation allows the incorporation of information about the reliability of measurements in terms of a weighting matrix. Alternatively, maximum likelihood and

maximum entropy methods can be used instead of GLS. To incorporate information from historical OD matrices, Tavana (2001) included a Bayesian inference that updates demand based on the results from the bi-level optimization. Alternatively, the distance between the estimated and target OD matrices could be incorporated into the objective function, as in Gupta (2005).

The upper level of the OD matrix estimation problem in Zhou (2004) is a weighted minimization of the deviation between the observed and simulated demand and link flows. Weights can be used in the upper level optimization function to reflect the level of reliability that the user wants to apply on demand or link flow measurements. Similar to Zhou (2004), Chi (2010) used adaptive weights on different components of the objective function. At the beginning of the estimation, higher weights were assigned to traffic measurements such as counts, speeds or travel times, since at the beginning of the process, these data are more reliable than the OD matrices from the demand model. As the system converges, a better estimation of OD is obtained, and the weight of the observed link counts is reduced in the optimization since they are not error-free. The adaptive weights can also mitigate the problem of over fitting of the observed counts. The optimum value of the weight can be obtained through least square estimation, or the model user can arbitrarily set them based on local knowledge. Another issue is that in congested networks, the volume is not an incremental function of demand; therefore, Chi (2010) proposed detecting congested segments temporally and spatially, and using density instead of volume in the objective function for congested segments, which is a better representative of traffic conditions.

Mahmassani et al. (2004) carried out the supply and demand estimation tasks in a sequential manner. They first calibrated the network as described in Section 2.2, and then used a bi-level optimization to estimate the OD matrices, similar to Tavana (2001). The authors investigated two different alternatives for the optimization part. The first approach was a linearly constrained GLS approach that minimizes the deviation between the estimated and observed link flows. The second approach was a weighted objective function whereby a higher weight was allocated to the links that carried more flow. In both approaches, weights were allocated to the objective function components, as discussed in Zhou (2004) and Chi (2010). The authors mentioned that using sparse matrix structure and decomposing the OD matrices into sequential sub-matrices can alleviate the problem of scalability. Fixing the OD cells that have no or little effect on traffic conditions and restarting the estimation with fewer variables increased efficiency.

Other approaches that were used to demand estimation are the Bayesian Inference and state-space framework, which are described below.

A simplified concept of Bayes Theorem is stated as Equation 2-4:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad OR \quad P(B|A) = \frac{P(B)P(A|B)}{P(A)} \quad (2-4)$$

Considering A as network conditions and B as travel behaviors, the formula above can be interpreted as predicting network (supply) behavior, given the demand ($P(A|B)$). Equivalently, it can be interpreted as predicting demand behavior, given the network conditions ($P(B|A)$). In the joint supply-demand calibration, the mutual relationship

between supply and travel behavior can be modeled through the Bayesian Inference, as shown in Figure 2-1.

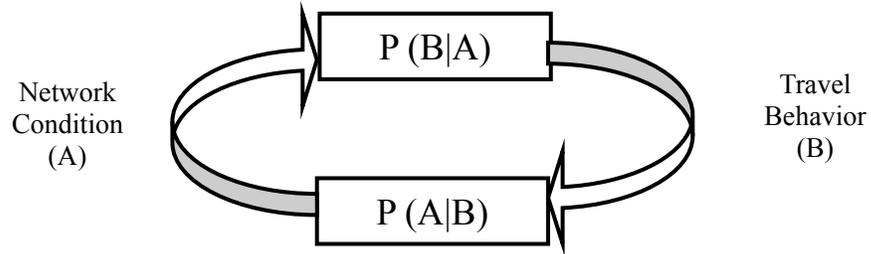


Figure 2-1 Interaction Between Demand and Supply

For a linear dynamic system, the state-space framework can be summarized, as shown in Equations 2-5 and 2-6 (Chen, 2003):

$$x_{h+1} = f * x_h + w_h \quad (2-5)$$

$$y_h = g * x_h + v_h \quad (2-6)$$

Equation 2-5 is called a “state or transition formula,” and shows how a state vector (x) evolves over time by evaluating $P(x_{h+1}|x_h)$, the probability of x_{h+1} , given x_h . The state vector can be OD flows, travel behavior parameters, speed-density relation parameters and so on. Equation 2-6 is called the “measurement equation” and maps the observation vector (y) to the unobserved state vector (x), or describes the probability $P(y_h|x_h)$. The model coefficients, f and g , need to be estimated, and w and v are model noises. The detector data, such as volume and speed, are examples of y . A well-known solution for the state-space model is Kalman filtering, in which model noises (w and v terms in Equations 2-5 and 2-6 are assumed to be a normal distribution with a mean of zero.

Ashok and Ben-Akiva (2002) and Lin (2006) modeled the relationship between demand and link flow as a state-space formula. It should be noted that in congested areas

with capacity constraint, link flows do not represent the demand. Capacity plus queue at the link can be used to approximate the demand. Hu and Chen (2004) estimated OD and travel time simultaneously through extended Kalman filtering. Zhou (2004) defined the true demand be estimated as a combination of regular pattern, structural deviation from the mean pattern, and random fluctuations, and applied Kalman filtering to capture these components. Kalman filtering was used as an external controller to inspect the adjusted OD before sending the OD estimation output to the DTA simulator.

The growth of ITS implementation is very promising in collecting full or partial trajectory data. With commercialized connected-vehicle devices mounted on cars, more trajectory information will be available in the future. Zijpp (1997) and Zhou (2004) were able to reduce OD estimation errors by combining AVI and count data. Dixon and Rilett (2002) deployed GLS and Kalman filtering to show the benefits of the incorporation of origin-destination and travel time information from AVI data.

Doblas and Benitez (2005) pointed out a practical aspect of OD estimation that was ignored in related studies. The preservation of the structure and pattern of initial OD should not be sacrificed to replicate traffic counts. Traffic counts reported by detectors are not error-free. Moreover, the information in the initial OD (usually from surveys or extensive calibration of travel demand forecasting models) is very valuable and expensive, and deviation from initial OD structure should be constrained. The authors modified the gradient-based algorithm of Spiess (1990b) implemented in commercial DTA tools to control the adjustment of the OD matrices by preserving the number of production and attraction trips for each zone. To optimally use the available data, Nguyen

(1982) incorporated production/attraction data from a historical OD matrix to a maximum entropy formulation.

In summary, the assignment-based OD matrix estimation problem that is of interest to this study was formulated using a number of methods, including bi-level optimization (utilizing a GLS or maximum likelihood approach), state-space problem, or Bayesian inference. The latter two methods can also be used in conjunction with the bi-level optimization problem to update the OD matrices based on the results from the optimization, in an iterative process. They can also be used as an external controller to limit the deviation of the estimated OD matrix from the initial or historical matrix. Depending on the source and quality of the initial or historical OD matrix, certain features of the matrix may be necessary to keep. For instance, some or all of the attraction production rates or some OD pairs might be kept constant during the estimation process.

2.5. Joint Supply and Demand Calibration

It is logical to suspect that there is a relationship between the supply calibration discussed in Section 2.3 and demand calibration discussed in Section 2.4. Supply calibration requires a good estimate of demand, and demand calibration requires a well-calibrated network. Doan (1999), Antoniou et al. (2007) and Vaze (2007) showed that joint supply-demand calibration is superior to the sole use of calibrating demand. There are two main approaches to demand-supply DTA calibration: sequential process that can be performed iteratively (Balakrishna, 2002; Mahmassani, 2004), or simultaneous estimation of all parameters (Balakrishna, 2007; Vaze, 2007).

Antoniou et al., (2007) utilized a nonlinear state-space model to jointly calibrate supply and demand in an online framework. Ashok and Ben-Akiva (1993) used the deviations of the model parameters from the best estimated parameter instead of the parameters themselves, as part of a joint supply-demand calibration process. This way, all available information (obtained from estimation in previous steps) would indirectly be incorporated into the model structure. The network was composed of 45 segments of a mainline freeway and associated ramps (no route choice behavior was involved). The author decomposed the problem and sequentially calibrated supply and demand parameters. Segment capacities were estimated according to the HCM methodology, and the TFM parameters were found by fitting the modified Greenshields model to sensor data for three grouped segments. Utilizing a similar approach, Vaze (2007) calibrated all network parameters, route choice parameters, and OD matrix elements in DynaMIT through state-space modeling, as well as through stochastic optimization modeling.

Chi (2010) conducted a weighted bi-level optimization to calibrate the supply parameters and estimate OD demands in a freeway system. The network (supply) was calibrated once before the OD estimation by fitting observed data to the modified Greenshields model, and once afterward to fine-tune the parameters obtained from the previous stage. Fine-tuning of the TFM parameters was carried out through bi-level optimization. The author also showed that the incorporation of an initial OD estimate can improve the overall performance of the estimation. In absence of historical OD estimates, a gravity model was used to produce an initial OD matrix. This matrix was then improved using a static OD estimation module that utilizes a maximum likelihood framework.

Balakrishna (2007) estimated all parameters of the supply and demand sides through stochastic optimization. Following Kunde (2000) and Vaze (2007), he utilized Simultaneous Perturbation Stochastic Approximation (SPSA) to simultaneously estimate hundreds of parameters on the network. Although this method is theoretically elegant, it has not been implemented successfully in real-world applications.

Interrelation between supply and demand was carried out through sequential and simultaneous processes. Simultaneous estimation of all parameters, although asserted to be more efficient, complicates the problem and limits the user's ability to monitor and control the change of parameters. Moreover, incorporating local knowledge about the network or the demand is difficult in this approach, since a large part of the optimization is automated. No successful application of this approach in the real world has been reported so far.

2.6. Convergence

Another issue that will be explored in this study is the quality of the traffic assignment solution, as measured by convergence. By definition, the user equilibrium is achieved when travelers cannot improve their travel times by selecting alternate paths, given their departure time. This implies that every used path between an origin and destination is a minimum cost path and that there are no changes in flow patterns or experienced travel times between assignment iterations after the convergence is approached. Convergence of the user equilibrium assignment is necessary to ensure the integrity of the resulting solution and to ensure that the model can be used in assessing alternative designs and operational strategies.

A number of approaches were proposed to solve the static and dynamic assignment problem. Some of these approaches are heuristic approaches, and others involve more rigorous mathematical programming (Ortúzar and Willumsen 2001). The mathematical programming approaches express the assignment problem as an objective function subject to constraints representing traffic flow properties.

The mathematical assignment methods generally allow the proof of optimality and uniqueness and produce superior solutions to those obtained utilizing the heuristic approaches. However, due to the complexity of the dynamic network loading functions required for DTA, the traffic flow models in DTA problems are generally non-differentiable. Therefore, heuristic algorithms that do not require derivative information are used for simulation-based DTA. Although with heuristic assignment, no formal convergence proof can be given, as is the case with mathematical solutions, measures of gap similar to those used in static equilibrium assignments that are based on mathematical solutions can be used to assess the quality of a solution. Still, heuristic approaches with simulation-based DTA fail to guarantee optimality and convergence.

Boyce et al. (2002) pointed out that a relative gap of 0.01% (0.0001) is required for static assignment so as to ensure sufficient convergence to achieve link-flow stability. There is no positive agreement on what represents an acceptable value of the relative gap in DTA. It was realized, however, that it is much more difficult to achieve a small relative gap in simulation-based DTA compared to static assignment, particularly for congested conditions (Chiu et al. 2011). The dynamic nature of traffic flow, particularly during congested conditions and the heuristic nature of the UE problem in DTA, makes it more difficult to achieve convergence in DTA, compared to STA.

A widely used measure for calculating convergence is called the “relative gap,” which measures the difference between the current iteration solution and the ideal solution. The ideal solution is loading the whole volume on the single shortest path (Chiu et al., 2011). This concept was applied with slight differences in the formulation in different studies. Link-based measures versus path-based measures have also been suggested by researchers, with recent discussions on the subject indicating that path-based (also referred to as trip-based) measures might be more meaningful (Chiu and Bustillos, 2009). Path-based or trip-based measures exploit disaggregate and tractable information of trips instead of aggregated link volumes. In addition, path-based criteria provide additional information that allows utilizing heuristics targeting those trips, travelers, households, or market segments that have the most impeding convergence to achieve better solutions (Resource Systems Group, 2010).

The relative gap should not be considered an ultimate qualification for the UE solution. A well-known problem of UE is that although it produces a unique set of link volumes, there can be multiple route solutions associated with these volumes. This can be a serious issue in problems such as select link analysis and subarea analysis. It is possible to define the unique desirable UE path set by setting some extra constraints on the assignment solution to avoid violating the conditions of stability and proportionality.

Bar-Gera et al. (2010) pointed out that even if the link flow reaches convergence, a main issue with route flows is that they are not uniquely determined by the UE conditions. Reaching path flow convergence is particularly important for applications, such as multi-class assignment, select link analysis, estimation of origin-destination flows from link flows, derivation of OD flows for a subarea of a

region, average travel time and average distance per OD in a generalized cost assignment, and so on. It was found that among all possible UE routes, there was just one that maximized the entropy, which should be considered the unique solution. It is proven that this solution also meets the proportionality condition. The proportionality requirement is defined by Bar Gera et al. (2010) in that the proportions of travelers on each of the two alternative segments should be the same regardless of their origin or their destination.

Lack of convergence can also affect the consistency and stability of the resulting solutions. Consistency is defined as the contribution of all eligible routes to the UE solution. This means that all routes should be included in the UE solution, unless there is a good reason for not being considered, like having a high generalized cost. Lu and Ni (2010) defined stability as the solution ability to accordingly respond to perturbation, meaning that if small changes in the network or demand are made, the model should respond to it with reasonable changes. On the other hand, Chiu and Bustillos (2009) and Peeta et al. (2011) state that a network is stable when link volume does not fluctuate, and a network is consistent when it responds appropriately to small perturbation.

A small relative gap does not assure a credible dynamic assignment solution. Lu and Ni (2010) showed that even with a very small relative gap (10^{-12}), misleading results that look reasonable may be obtained, yet respond unreasonably to small perturbation. For instance, a 10% decrease in capacity of a secondary road might cause serious congestion in another part of the network. Consistency, proportionality, and stability are needed to check for the evaluation of alternative treatments of the transportation system, and for applying methods such as select link analysis, select zone analysis, and subarea analysis. This is also very important to ensure unique solutions of multi-class

assignments, particularly in ML where preferential treatments of some of the classes are applied (Boyce et al. 2010).

2.7. Summary

Managed lanes are accepted as effective countermeasures against freeway congestion. These facilities are proactively operated in response to traffic situations, by means of access management, variable toll policies, and vehicle eligibility constraints. Assignment is a critical step in ML demand forecasting to determine the effect of pricing on drivers' route choice behaviors. DTA, combined with mesoscopic or microscopic simulation is identified as an effective tool for ML assignment. DTA models are more sensitive to the level of congestion and temporal demand/network changes, and compared to STA, more realistically model the route choice behavior.

A model can only replicate real-world observations when supply, demand, and route choice calibrations are completed. Network or supply calibration entails estimating capacity and traffic flow model parameters for each link in the network. These parameters affect the travel time, congestion time, queue formation and queue spillback when the demand is loaded. Demand calibration is used to estimate a trip table that produces observed link counts and congestion patterns when loaded onto the network. Route choice calibration involves the selection of the parameters and methods of the assignment process.

Joint calibration of network, demand, and route choice parameters is confirmed to be superior to separately calibrating these components. Two different approaches, sequential and simultaneous calibration, have been used by researchers and practitioners.

Despite the elegance of the mathematical formulations and solutions for simultaneous network and demand calibration, their implementations in the real world are not straightforward and have not been executed. Once the network, demand, and route choice parameters are selected as described above, additional fine-tuning of the parameter may be needed to adjust local variables to produce the observed queues and operations

It should be noted that replicating traffic volumes does not guarantee a well-calibrated network. Temporal-spatial congestion patterns should be reasonably replicated. Estimated OD matrices should also be consistent with other sources of data, such as zonal information from the production/attraction step or from the trip distribution step and certain attributes of the historical OD matrices. Simulated queue length and/or density are other measures that should be checked against the estimated values from field observations when the network is congested and the demand is not easy or possible to obtain. In the objective function used to estimate OD matrices, adjustable weights on different components can reflect the level of confidence in the data and improve the performance of the estimation. These weights can also reflect the importance of individual segments of interests, such as bottlenecks or locations with volumes that better replicate the changes in demand patterns.

Ranking links based on their contribution in updating OD routes reduces computational time. Also, OD elements that do not significantly affect the assignment can be fixed to reduce the size of the OD estimation problem. OD matrices can be aggregated into longer time segments, compared to the observed data time interval, so as to alleviate the problem of under-determinacy. Origin-destination survey data is very valuable if available, and a structural deviation from it should be avoided. Different logic

and reasonableness criteria should be devised into the OD estimation procedure as a feedback process to avoid error propagation.

Different methods of OD estimation should be empirically tested to determine which method can better preserve the historical OD pattern, which is the most computationally efficient, and which can better replicate congested network conditions. Investigating the optimal modeling of the supply-demand joint calibration also requires empirical testing. Assignment convergence and joint calibration convergence should be properly addressed and checked. Convergence should be checked for each time interval and for each OD pairs.

3. METHODOLOGY

This chapter presents an overview of the methodology and tasks implemented throughout this study. This chapter also explains the task sequence and the link between different tasks. More detailed descriptions can be found in the subsequent chapters.

Figure 3-1 shows a schematic summary of these tasks and the interrelations between them. Task boxes and feedback loops are color coded for better understanding of the relation between different components. The developed ML modeling framework starts with acquisition, validation, and the processing of data from multiple sources. Initial network and demand data are obtained from the regional demand forecasting model. Then, network geometry is updated based on aerial photograph from Google Earth. Data are also obtained from microwave detectors, Portable Traffic Monitoring Sites (PTMS), Telemetered Traffic Monitoring Sites (TTMS), and managed lane tolling systems. . Additional demand information is extracted from a previous microscopic simulation study. Data pre-processing procedures are developed and implemented in this study, as detailed in Chapter 4. These procedures include the identification of representative days (by filtering out incident days, special events, weekends, and applying classification techniques for better filtering), data fusion and aggregation, data validation and cleaning, and checking detector data for spatial and temporal consistency.

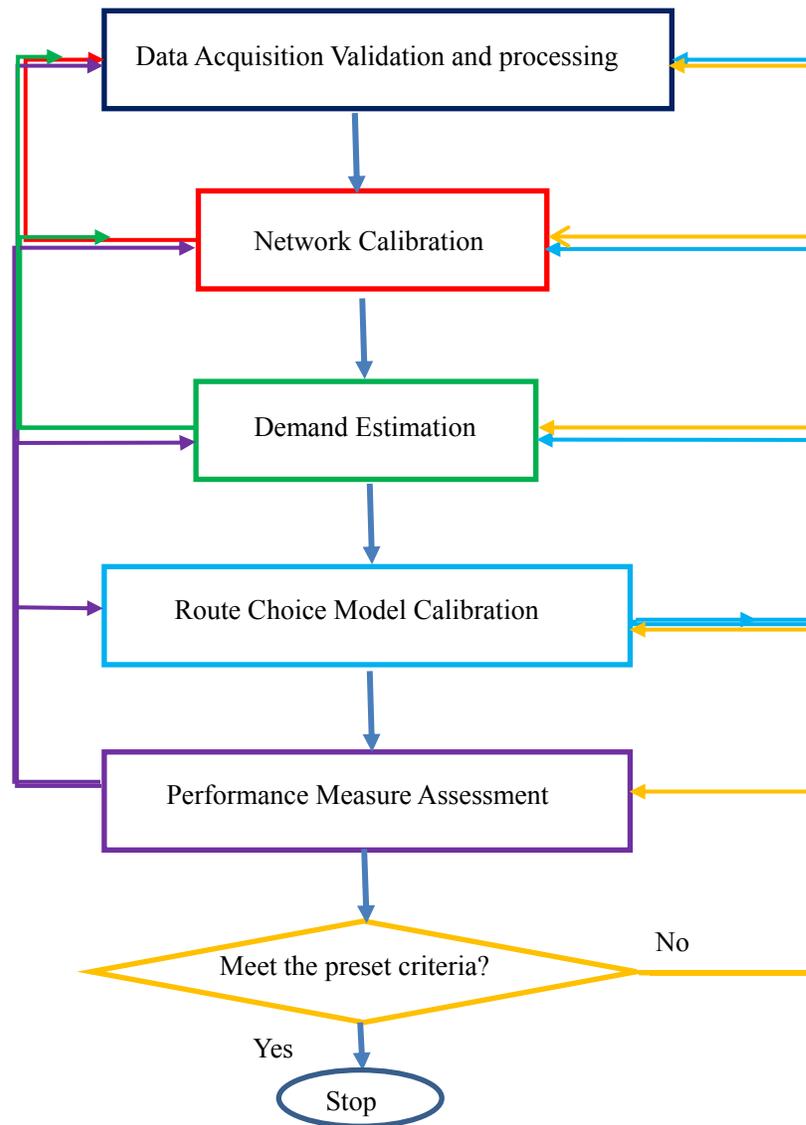


Figure 3-1. Methodology Flowchart

The next step is network calibration (also referred to in this study as supply calibration). Network calibration consists of estimating link capacity and traffic flow model parameters based on data collected at bottleneck locations, adjusting traffic management and control parameters, and fine tuning the calibration parameters at other

locations of the network to replicate real-world traffic conditions. The network calibration sub-tasks such as developing and using methods for bottleneck identification, free flow speed and capacity estimation based on field measurement, and traffic flow model parameters estimation are describe in detail in Chapter 5. Since detector speed data are available for each lane and for intervals as short as one minute, free flow speed can be measured in uncongested intervals. The availability of speed and count data also allows for the estimation of capacity based on field data. As was described in Chapter 2, researchers have proposed different estimations of capacity, either as a deterministic value, or a probabilistic distribution. The present study estimated capacity at bottleneck based on deterministic definitions, and these values have been compared to HCM recommended values. Running assignment models (both static and dynamic) with HCM estimated capacity and field estimated capacity demonstrates the significant effect of capacity coding on congestion and route choice behavior. At the beginning, the network is loaded with the initial demand taken from a regional planning model. After the initial estimation of network parameters, an OD (demand) estimation process can be run to improve the initial demand

The purpose of demand estimation is to improve the initial demand extracted from the regional model for use in the DTA model. The first task is to distribute the 3-hour regional demand over a 15-minute DTA interval based on observed variation in link volumes. However, the result of factorization, when loaded onto the network, fails to produce real-world traffic measures. Thus, a procedure called static OD estimation is performed to estimate new OD trips that, when loaded on the network, can produce results that better replicate measures such as link volume and speed. At the same time,

the utilized procedure ensures that a significant, unjustifiable deviation from initial trip tables is avoided. The static OD matrix estimation process is implemented using the Cube Analyst program, a tool that estimates trip matrices based on the maximum likelihood technique, coupled with an optimization procedure. The tool utilizes data from different sources and considers different levels of confidence or reliability inputted by the user for each source of data. Not only can the data include traffic counts and prior (seed) matrices, but also partially observed matrices, zonal trip end (generation and attraction) data, vehicle routing, travel cost matrices, and even previously calibrated trip cost distribution functions.

Vehicle routing information is a very important input for estimating ODs and is produced by the traffic assignment tool. In static OD estimation, as the name implies, the utilized traffic assignment is STA. In more advanced dynamic OD estimation however, the utilized assignment tool is DTA. The drawback of static OD estimation is that static traffic assignment cannot properly capture the congestion and significant delay due to queue formation and spillback. Therefore, the result of this procedure tends to underestimate ODs. For this reason, dynamic OD estimation was initially tried in this study. However, due to software limitations and immaturity, this approach was not used. To circumvent the aforementioned problem, queue, as a measure of congestion, is calculated on each screenline over time, and is added to the screenline volume that Analyst aims to replicate. Further discussion of the demand estimation procedure is presented in Chapter 6. This demand should be fed into the previous step of network calibration, for a better estimation of network parameter with more accurate demand.

The last task, described in detail in Chapter 7, is to model and calibrate the route choice behavior in the assignment step. Up to this point, the default parameters of the route choice model were used to estimate network properties and OD trips. With a better calibrated network and demand, assignment parameters can now be estimated. The result of this step, significantly affects demand estimation, and consequently, the network calibration. Therefore these procedures need to be carried out another time, and the entire procedure shown in Figure 3-1 is repeated until desired convergence is achieved.

Route choice behavior is modeled using two different approaches. In the first approach, the toll cost is converted to the equivalent travel time and is added to the link generalized cost function. In the second approach, prior to the traffic assignment, a willingness-to-pay curve is used to determine the percentage of travelers who are not willing to pay the toll. In the assignment step, this group of users is prohibited from using ML. The remainder of the assignment is governed by user equilibrium, based on the travel time on ML and GPL. In both approaches, the toll is dynamically updated every 15 minutes, based on the maximum density on ML. Different parameters such as toll schedule, value of time, and the willingness-to-pay curve have been calibrated using multiple-source data. These two approaches are compared based on different performance measures, such as replication of the diversion to ML, convergence, and stability of the assignment solution. The effectiveness of DTA versus STA in ML modeling is also evaluated on the real-world network.

For each of the aforementioned tasks, performance measures have been identified and evaluated. It is worth emphasizing the necessity and benefits of feedback loops

between all tasks. Examples of the role of the feedback are presented in the following chapters.

4. DATA ACQUISITION AND VALIDATION

Advanced modeling tools, such as DTA, demand more detailed and higher-quality data to ensure that the developed model accurately replicates real-world conditions. Compared to STA models, DTA requires more refined network representation and additional data details, both temporally and spatially. Moreover, congestion data such as queue presence and queue length should be incorporated into DTA calibration, while such data is generally not used in STA-based tools. Traffic control and management details are also needed if the impacts of traffic control and management are to be accurately modeled.

In this study, the network and an initial estimation of the associated trips were extracted from a regional planning model. The performed network editing efforts and refinement of the initial demand for use in DTA are discussed in this section. This chapter also describes the collection of traffic detector data that provides estimates of measures, which are essential to the development and calibration of simulation-based DTA tool applications. Detector data requires careful examination and a significant amount of time for filtering and processing to exclude and/or correct suspect data.

4.1. Network and Demand Data Extraction

4.1.1. Subarea Network and Matrix Extraction

The study area was extracted as a subarea network from the SERPM model validated in 2010, as depicted in Figure 4-1. The subarea boundary can be specified using the Cube Polygon feature or a GIS tool. The Cube can then be used to extract the subarea

network from the SERPM model network by using this predefined subarea boundary. The results of this extraction are a subarea network and associated trip tables for multiple users.

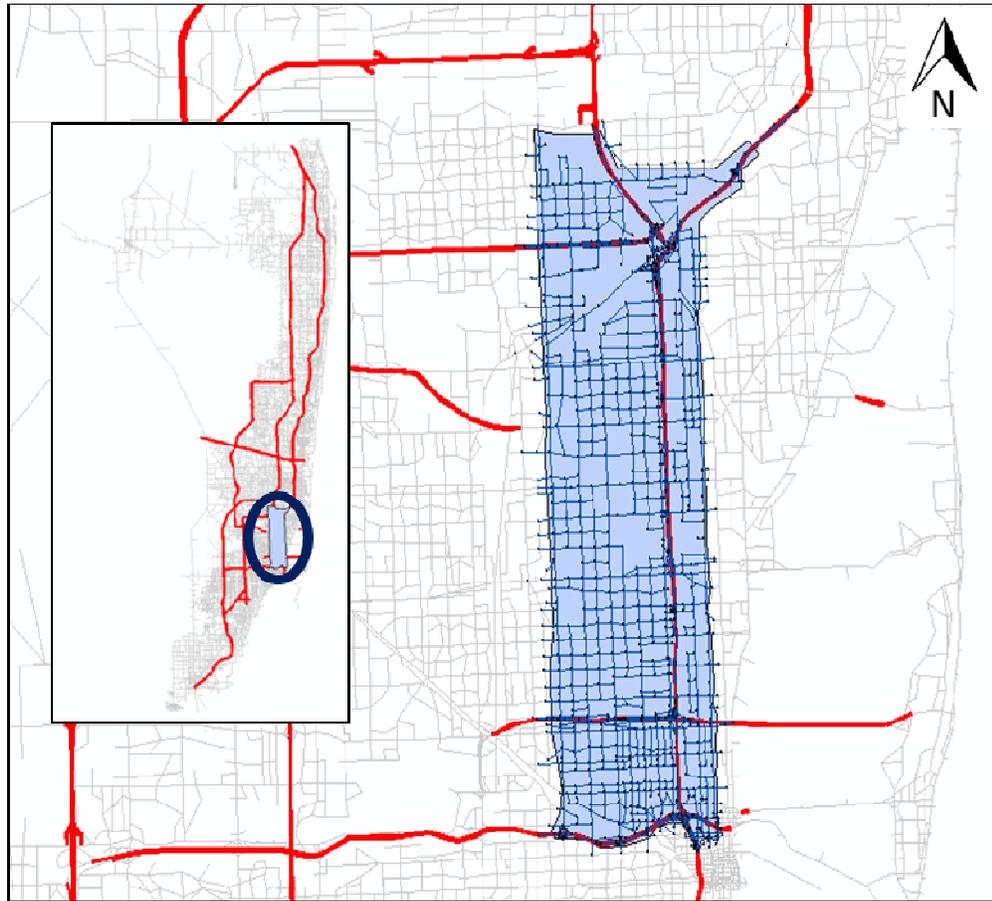


Figure 4-1 The Extracted Subarea from the SERPM Model

The extracted subarea contains new node and zone numbers and the Cube stores the association between the old numbers (in the original network) and the numbers in the new network (in the subtracted network).

The network geometry needs to be updated to better represent the existing real-world network since the details and accuracy of the network in demand forecasting models are not sufficient for DTA applications. The attributes of each link is adjusted in

this study based on the Google Earth map. The network geometry update is performed following the procedure presented in Figure 4-2. The subarea network is converted into the KML format for the Google Earth application and the SHP file format for the ArcGIS application.

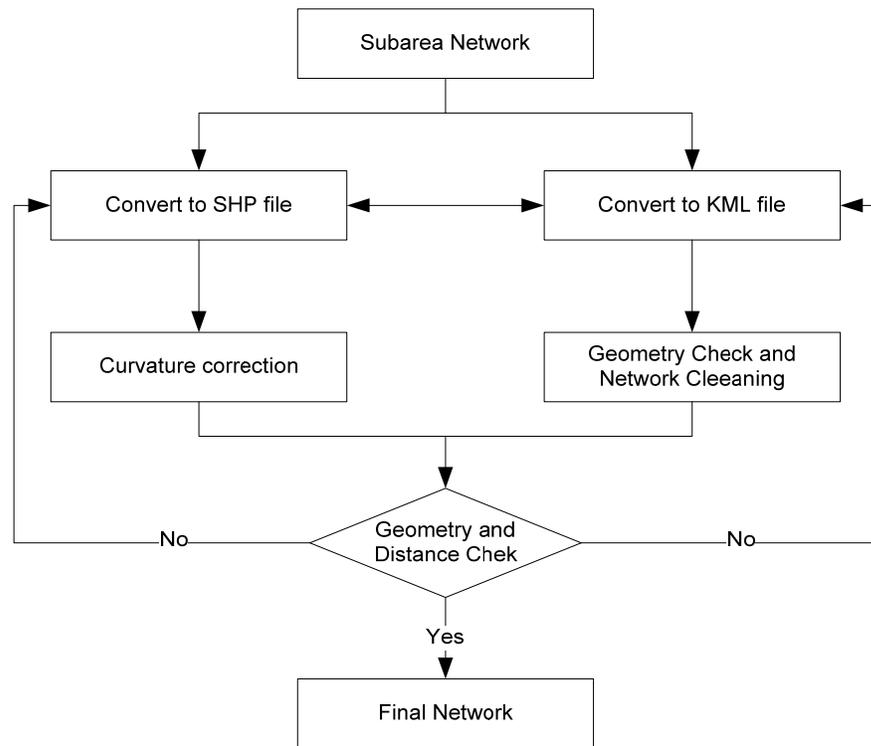


Figure 4-2 Network Geometry and Distance Update Procedure Flowchart

Imposing the network on the Google Earth map allows correcting the network curvatures, connections, and other geometry attributes. The link lengths are accordingly modified. The links in the original network file in the demand forecasting model are established based on direct node-to-node connections. Therefore, all of the links in the demand forecasting model are straight lines. In order to obtain the real-world curvature of the links, the network was converted into the SHP file format from the Cube network format, allowing the links curvature to be drawn based on the real curvature using the

GIS modification tool. All of the links' lengths were updated based on the identified curvatures. Based on prior experience with the DTA tool used in this study (Cube Avenue), short links can produce unrealistic congestion. Therefore, it is very important to identify these links in the extracted subarea network and properly adjust their lengths to prevent the unrealistic congestion from occurring. It was found that in most cases, the issue of short links could be addressed by updating the links' lengths, considering the true curvature of the links, and moving the merge/diverge nodes based on their real-world location in Google Earth maps. Figure 4-3 shows an example of the network geometry adjustment conducted as part of this study.

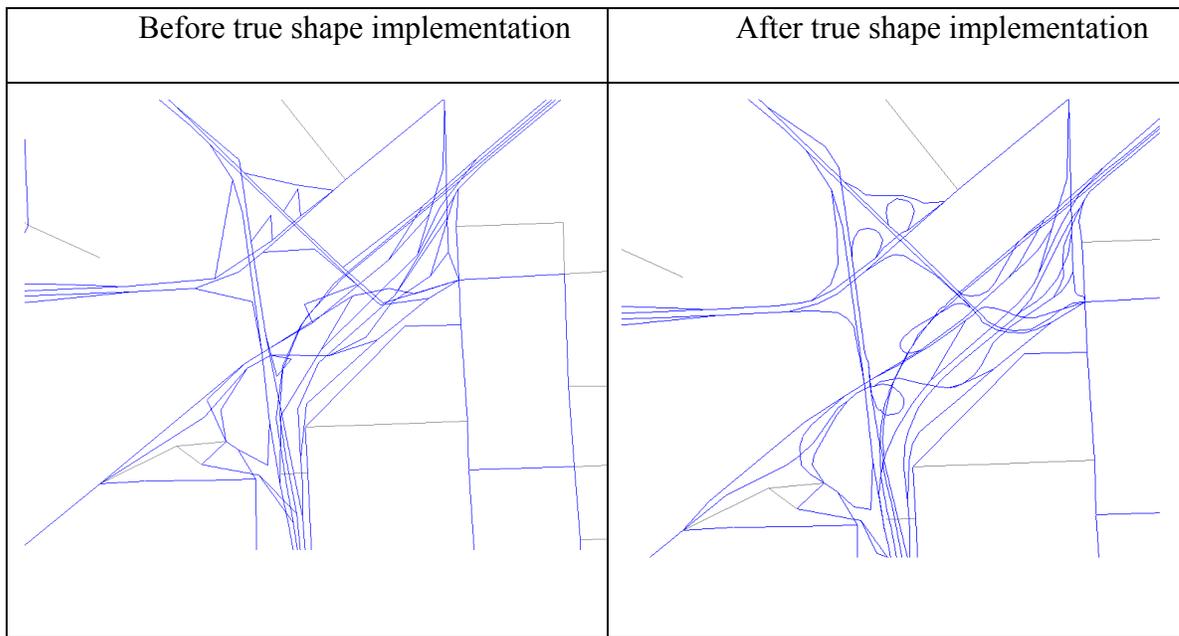


Figure 4-3 Network Curvature Correction

Modifying zones and connectors may also be necessary and should be considered. Another important consideration in the cleaning process is checking the consistency of the number of lanes between successive links, especially in merge and diverge segments, and at intersections with exclusive left- and right-turning lanes.

Most of the work conducted in this study was performed on a linear North-South corridor shown in Figure 4-1 that represents the I-95, and includes ML, GPL, and associated on- and off-ramps. In most segments, the freeway includes four general purpose lanes and two managed lanes, which are separated from one another by a soft barrier. This corridor contains 57 zones, 303 nodes, 303 links, and a total demand of 117,541 vehicles for three hour period.

4.1.2. Demand Data

In this study, initial trip matrices were extracted from the SERPM regional demand model. Regional travel demand models represent an important source of OD trip information which is estimated through detailed and approved processes that ensure consistent behaviors of travelers in the demand generation, distribution, and mode choice steps. However, some issues with these models include lack of detailed model calibration at the subarea level and the potential changes in the network and demands since the model's last calibration. Even more critical to DTA modeling is that the regional demands are forecast for daily trips or three to four hours of time-of-day model period. These demands need to be distributed over shorter time intervals for DTA applications. The most common interval study for DTA modeling is 15 minutes.

4.2. Detector Data Acquisition and Preprocessing

Detector data collected by the Florida Department of Transportation (FDOT) District 6 Traffic Management Center (TMC) is extensively used for the demand estimation, model calibration, and validation. The corridor of interest is instrumented every 0.3 to 0.5 mile with microwave detectors that report volume, speed, and density

measurements in 20-second intervals for each lane. This data were obtained from the Statewide Transportation Engineering Warehouse for Archived Regional Data (STEWARD). The STEWARD database contains summaries of traffic volumes, speeds, occupancies, and travel times aggregated by 5-, 15-, and 60-minute periods, as requested by the user. Using a Web-based interface, the user can specify date and time ranges and detector locations for which the data are needed. This data is supplemented by measurements from PTMS ramp counts from the FDOT Statistics Office. The PTMS data include 15-minute ramp counts for two or three days per year. No speed or classification data is available. Ramp counts obtained from the PTMS and ramp metering detectors represent the total origin and destination demand on the linear network and are very useful in the demand estimation process.

Table 4-1 lists the numbers of the available microwave and PTMS detectors in the corridor network and the selected detectors after removing redundant or erroneous detectors.

Table 4-1 Available and Selected Detectors

Detector Station	No. of Stations	
	Available	Selected
ITS	109	87
General Purpose Lane	78	56
Express Lane	31	31
PTMS	150	150
Mainline	10	10
Ramps	99	99

By imposing the network and detector maps onto Google Earth's map, it was possible to manually associate the detectors in Table 4-1 with network links. If any link is associated with more than one detector, only the most reliable one was kept.

Truck percentages are available for ramps from PTMS data. For the mainline, the truck percentage was obtained from nearby permanent TTMS stations, also operated by the FDOT Statistics Office. These percentages were confirmed by manual counting of recorded videos at selected corridor locations.

4.3. Other Data Sources

Data from other sources were also obtained and used in this study, as listed below:

- A previously calibrated micro-simulation model of the study area that includes traffic demand estimates
- Real-world ML toll values for each 15-minute interval from FDOT District 6 TMC
- Ramp metering data from FDOT District 6 TMC
- A previously calibrated logit model for ML willingness-to-pay prediction along the I-95 corridor

4.4. Data Preprocessing and Validation

Data pre-processing procedures include the identification of representative days by filtering out incident days, special events, weekends, and applying classification techniques for better filtering, data fusion and aggregation, data validation and cleaning, and spatial and temporal consistency checking of detector data.

Demand and congestion patterns vary greatly day by day. The representative days for modeling and calibration are considered to be weekdays (Tuesday through Thursday)

without incidents or abnormal external conditions such as heavy rain. Non-representative days can be filtered out by different methods that exclude days with special events or conditions. Also, data mining methods can exclude days with significantly different volumes or speed patterns from normal days.

Between May 2010 and May 2011, 16 days were identified based on detector data as ideal days to represent normal day traffic. Among these days, the speed varies with a coefficient of variance between 5% to 20% for different detector locations, and the volume varies with a coefficient variance between 3% to 7% for different locations. For different purposes, a specific day or an average of all representative days may be used for calibration. Using the median day may be better than using the averages, since the averages do not represent any of the real-world days.

Inconsistency between consecutive detector counts is a major consideration. Sometimes it is not enough to compare just one pair of detectors, and there is a need to check several stations upstream and downstream of each location. The addition of on-ramp volume and subtraction of off-ramp volume to estimate the expected volume for the station can be used as a reference to assess the accuracy of the measurements. In the presence of queue, this procedure becomes more complicated, and the capacity constraints should be considered. Figure 4-4 is an example of two successive detectors with an on-ramp between them, with approximately 190 vehicles per 15-minute intervals. The upstream and downstream detectors, however, show the exact number of counts. It should be noted that the reported counts are below capacity at all times, therefore, this issue is not caused by capacity restrictions. The comparison of detectors with additional

upstream and downstream detectors disclosed that the detector located downstream (Detector 7) is not reliable.

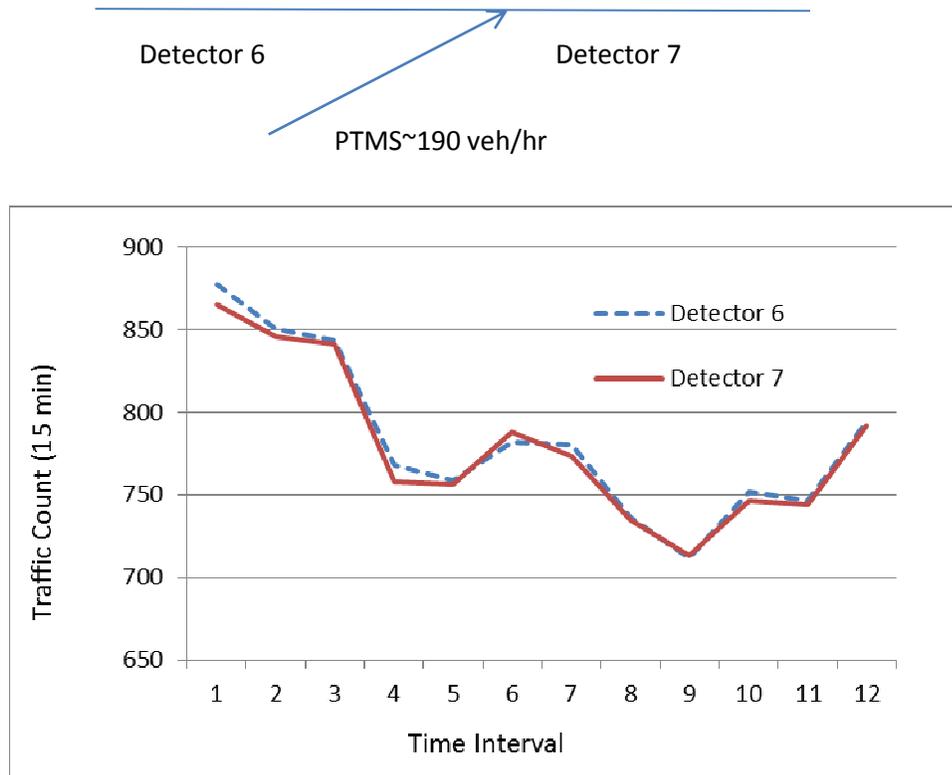


Figure 4-4 Volume Inconsistency Between Successive Detectors

Figure 4-5 shows another example of volume inconsistency between successive detectors. Selecting the right detector for each segment is only possible by having benchmarks, reliable detectors upstream and downstream of the segment, and selecting the most reliable detector by calculating the volumes from several upstream/downstream detectors, as previously discussed.

In addition, the number of lanes that the detector covers, which is included as an attribute in the detector database, should be checked, because some detector counts

include mainline and merge/diverge volumes. The consistency between detector count and estimated link capacity should also be checked to ensure that the reported count is below the segment capacity.

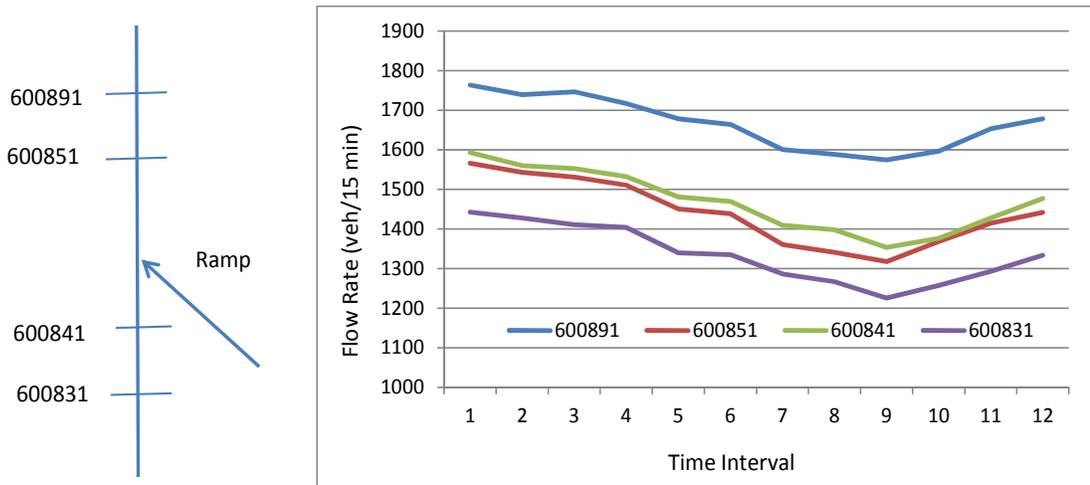


Figure 4-5 Volume Inconsistency Between Successive Detectors

With advances in traffic surveillance technologies, collecting, archiving, and analyzing traffic data is becoming more accessible and affordable. Now is the time for agencies to make the most out of these data for successful demand estimation, network calibration, and consequently, successful operation of managed lane facilities. The present study shows how data from multiple sources can be integrated, validated, and best used in different stages of modeling and calibrating. Extensive and careful processing of demand, traffic, and toll data, as well as proper definition of performance measures, resulted in a calibrated and stable model, which closely replicates real-world congestion pattern, and can reasonably respond to perturbations in network and demand properties.

Whenever data is available, comparing PTMS and ITS microwave detector count data may improve the reliability of the data. In this study, it was found that there is an

acceptable match between ITS and PTMS counts on the ramps. On the mainline, however, PTMS reported higher volumes, compared to ITS data in the PM peak. Manual counts of recorded videos were conducted to validate the data. It was found that the manual counts are closer to ITS data than the PTMS. Figure 4-6 to

Figure 4-8 show the comparison of PTMS and ITS volume data for three days: August 9, August 10 and August 11, 2011.

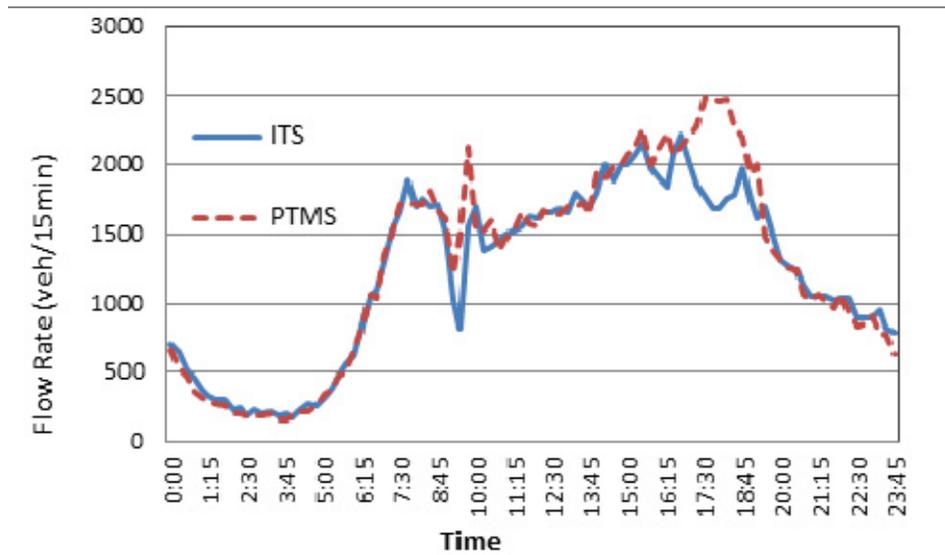


Figure 4-6 Comparison of PTMS vs. ITS Volume Data (August 9, 2011)

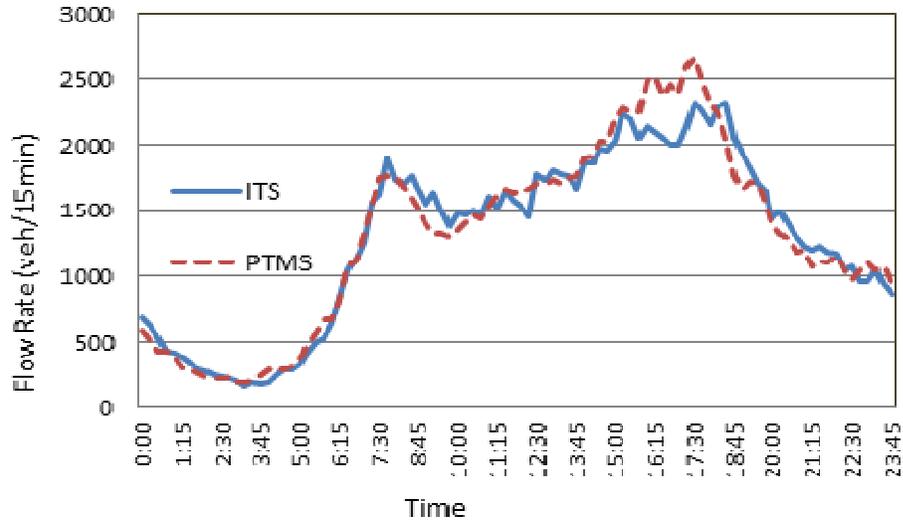


Figure 4-7 Comparison of PTMS vs. ITS Volume Data (August 10, 2011)

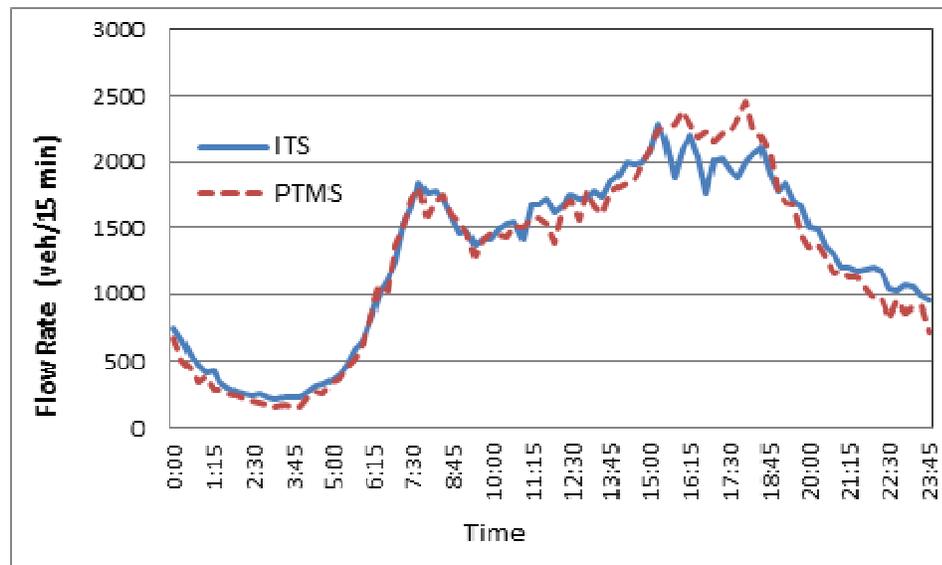


Figure 4-8 Comparison of PTMS vs ITS Volume Data (August 11, 2011)

ITS data normally do not include detectors for the on- and off-ramp locations, unless ramp metering exists. The study section includes ramp metering and thus, ramp detectors. In terms of ramp metering, there are three types of detectors: upstream (queue) detectors that measure the demand, and downstream (arrival and departure) detectors

before and after ramp signals. When ramps are equipped with ramp metering, the modeler should decide which information to use: either the upstream demand, or the volume that passes through the ramp metering. For OD estimation purposes, the former should be used.

Detailed examination of the ITS data may help to identify the reason(s) for the congestion, so as to assist in the calibration process. Figure 4-9 and Figure 4-10 show lane by lane data of speed and occupancy for one detector at a congested location. This detector location was initially defined as an active bottleneck for potential capacity measurement. Lane by lane data of speed and occupancy, however, revealed that the congestion at this location is caused by a spillback from a downstream off-ramp. Therefore, the two left lanes have considerably lower speeds and higher occupancy than the other lanes, indicating that this location is not a candidate for use in estimating capacity.

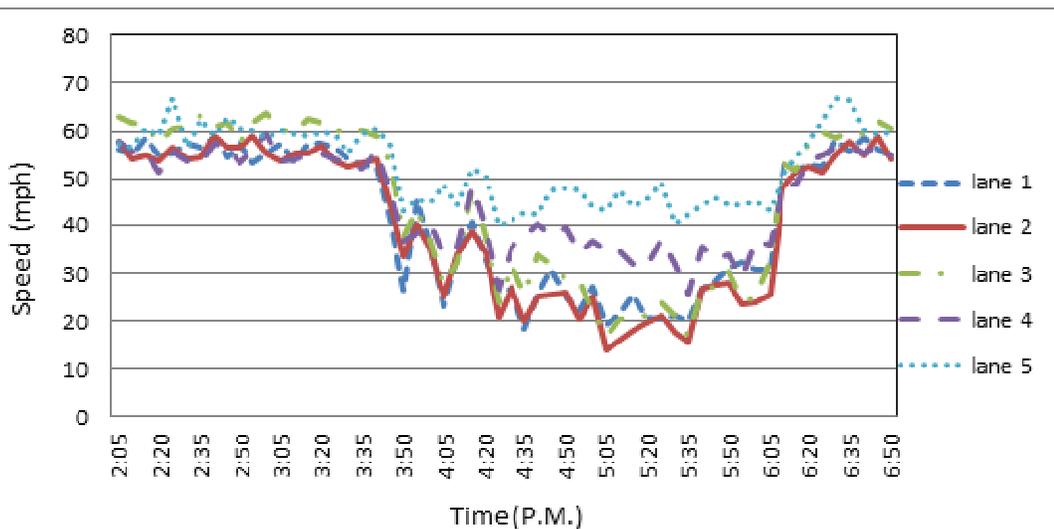


Figure 4-9 Lane by Lane Speed Data

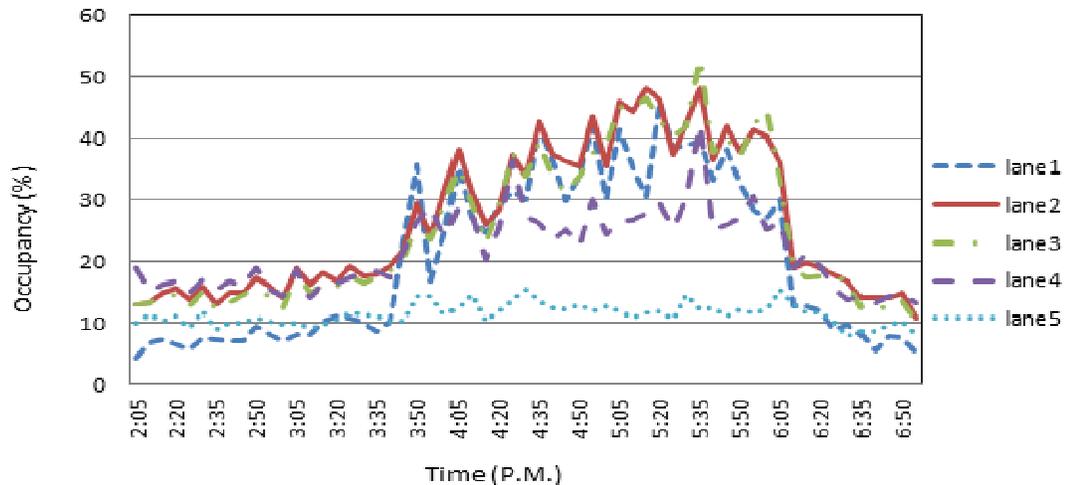


Figure 4-10 Lane by Lane Occupancy Data

During the OD estimation process, it was found that the detector for one of the screenlines does not produce correct volumes. It should be noted that this value passed through all abovementioned filtering processes. Utilizing this screenline data significantly affected the OD estimation process. This example shows that comparing the data from multiple sources of information should be a continuous and iterative process throughout the modeling and calibration tasks.

Depending on the network under consideration, there may be a need to disaggregate the zones from larger zones used in the regional model to smaller zones. There may also be a need to modify the zone connector setting. Careful examination is needed to understand how the zones and their connection setup affect the results of the modeling.

4.5. Summary

With advances in traffic surveillance technologies, collecting, archiving, and analyzing traffic data is becoming more efficient and affordable. However, these data

have not been effectively used in transportation system modeling. This study collected data from different sources to successfully calibrate a complicated route choice model with managed lane modeling. The present study integrated, validated, and effectively used data from multiple sources in different stages of ML modeling and calibration. Extensive effort was dedicated to data validation and pre-processing.

5. SUPPLY CALIBRATION

Supply or network calibration estimates the network parameters such as capacity and traffic flow model parameters that define network performance in producing travel time, forming queues, and queue spillback. From literature review, a systematic multilevel approach of network calibration is adopted in this study, with an increasing calibration scope in each level. The process starts at the level of separated bottlenecks where capacity is estimated by various methods based on field data. The network is gradually extended to connect the bottlenecks and then to the whole corridor and subarea coverage.

The advantage of this approach is twofold: First, critical spots of the network can be better identified, analyzed and replicated. Second, a more reliable demand can be estimated for the smaller networks that are the focus of this study, which is very important in the iterative process of demand-network calibration. Focusing on isolated bottleneck locations and the freeway corridor for managed lane assessment enables the capturing of the interactions between supply and demand in addressing the causes for congestion, which is not tractable in more complicated networks.

Speed time-space contours are used extensively as part of the methodology of this study to identify traffic and bottleneck conditions, and their impacts. Figure 5-1 displays speed contours for representative days of low, medium, and high congestion, and an average for all selected days. As can be seen in this figure, the traffic patterns and the reason for the congestion can vary from day-to-day, even after removing non-representative days.

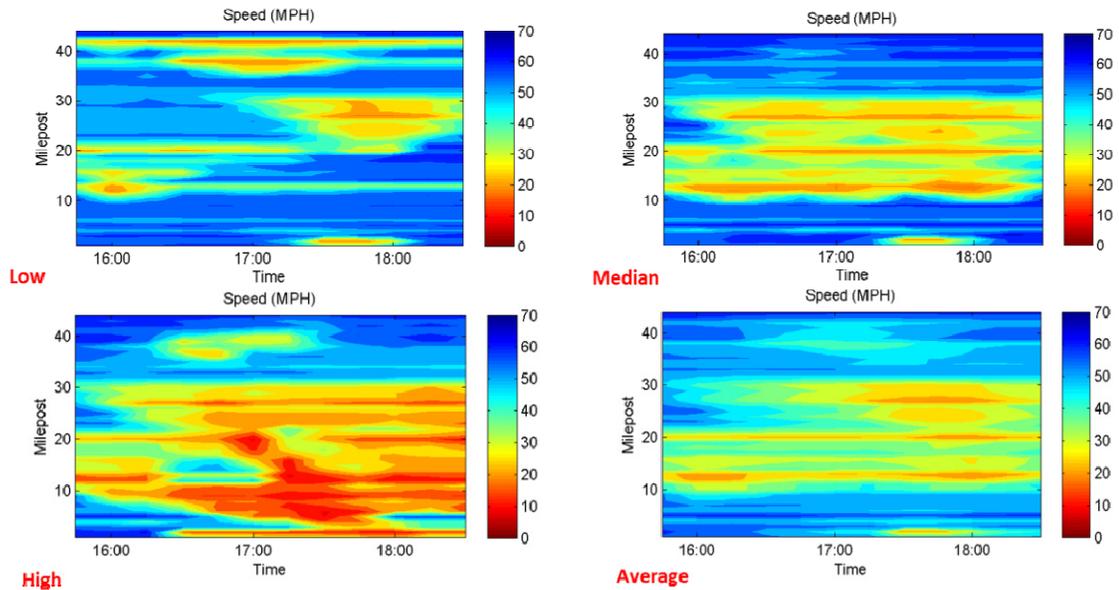


Figure 5-1 Speed Contour for Different Classes of Demand

In this research, initial demand matrices were obtained utilizing the network subtraction process from a regional travel demand forecasting model. Regional travel demand models represent a very important source of OD information that is consistent with the behavior of travelers, as modeled in demand generation, distribution, and mode choice steps.

The regional matrix covers the entire study period of three hours. Acquiring time-dependent trip tables at 15-minute intervals that reflect the current demand and traffic situations required the use of a sequential scheme that iterates between the supply and demand calibrations until convergence. The details of the demand estimation process are presented in Chapter 6.

It should be mentioned here that before the start of the calibration process, checking for mistakes in coding was conducted to omit any errors. In addition, the

model “validity” was checked according to FHWA guidance (Sloboden et al., 2012) including conducting a series of stress tests and diagnostic testing steps.

5.1. Bottleneck Identification

In this study, visualization techniques, in combination with comparisons between upstream and downstream measures, were used to identify congested areas and bottleneck locations. Based on the speed contours presented in Figure 5-1, Stations 12, 20, and 28 were initially identified as bottleneck locations in the PM peak period. Stations 12 and 20 are located in the on-ramp merging areas after the acceleration lane drops. Lane-by-lane data analysis of the ITS detector located at Station 28, however, showed that the congestion in this location is definitely caused by a backup from an off-ramp exit to a major freeway (the Florida Turnpike), causing low speeds and high occupancy in the two left lanes, while the three right lanes have light congestion. Thus, the only bottleneck locations that can be used to estimate capacity are those at Stations 12 and 20.

5.2. Free-Flow Speed

In the network under study, based on the Highway Capacity Manual (HCM) 2000, assuming a 6-foot lateral clearance, for a lane width of 11 feet and interchange density of 1.16, the Free-Flow Speed (FFS) is estimated to be 63 mph for segments with three lanes, and 64 mph for segments with four lanes. Based on the HCM 2010 analysis, the FFS is estimated to be around 66.9 mph for most segments (FFS is not depending on the number of lanes in HCM 2010). Based on a combined criterion of volume less than 1000 pc/hr/ln,

and occupancy below 10 percent, the FFS values were derived from detector data. Estimating FFS as the 85th percentile of speed over several days as suggested in literature showed very similar results. This value greatly varies between stations (from 54 mph to 64 mph), with an average of 59 mph, which is significantly lower than the HCM 2000 and particularly, the HCM 2010 estimates, as shown in Figure 4-2. It is worth mentioning that the posted speed on all I-95 corridor segment studied is 55 mph.

A previous study on an adjacent corridor (Florida State Road 826) with the same speed limit shows similar differences between the values estimated by the HCM 2000 and HCM 2010; however, it shows a higher measured FFS compared to the present study (Xiao et al., 2010). It is expected that the selected I-95 segment operates differently from an average corridor since it passes through a dense urban environment with frequent interchanges, has vertical and horizontal alignments that may affect capacity, and includes parallel managed lanes that are separated from the general use lanes by soft barriers.

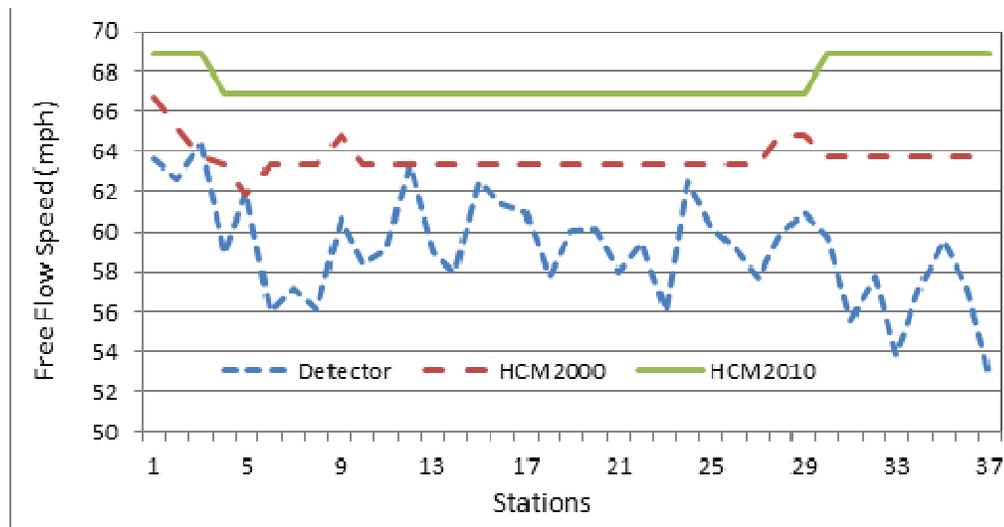


Figure 5-2 Variation of the FFS Along the Corridor (I-95 NB)

5.3. Capacity Estimation

This section presents a comparison between the capacity values estimated based on different sources and utilizing different methods. HCM is the primary source for estimating highway capacity for planning and operation applications. The HCM capacity values are expressed in personal car per lane per hour and should be converted to vehicle per lane per hour by considering heavy vehicle percentage for comparison with real-world measurements. The heavy vehicle percentage was estimated to be around 5%, based on recorded video observations. The HCM provides adjustment factors for different weather conditions and the degree of familiarity of the drivers with the road (driving population factor). The selected representative days of this study included normal weather conditions, and no necessary adjustments. The driver population factor has a significant effect on adjusting capacity, but is very difficult to obtain, and there is no guideline in the HCM on how to estimate it. The HCM mentions that this value usually varies between 0.85 and 1, and recommends using 1, unless there is sufficient evidence to reduce it, though a default value of 0.95 is mentioned for urban freeways. As is shown in Table 4-1, the values coded in the Southeast Regional Planning Model (SERPM) and estimated based on the Florida LOS/QS manual (5,6) corresponds to those values estimated by HCM for the 5% for percentage of trucks, and 95% for the familiar driver population.

To reflect site specifications, the capacity was also estimated based on detector volume data, aggregated at 15-minute intervals according to the HCM definition of capacity. In order to ensure that the only data utilized in estimating capacity at the

bottlenecks are for intervals not affected by downstream congestion, an examination of speed contours was made so as to identify and exclude intervals in which the capacities of the bottlenecks are affected by a spillback from downstream. The difference that resulted from the removal of the data from these intervals in capacity measurements for some methods is presented in Table 5-1. The results clearly show the need for this step. For example, the capacity measurement based on the Rakha method is 1,710 vph without removing the spillback intervals and 1,800 vph when the data from these intervals are removed. Table 5-1 indicates, based on different methods including the pre-breakdown flow method, the Rakha model-based method, and the maximum occupancy method, that the capacity before breakdown is about 1,850 vph. The queue discharge rate appears to be lower than this value based on the results in Table 5-1.

Table 5-1 Estimated Capacity at Active Bottleneck Locations (VPH)

Method	Station	Station	Reference
	600561	600711	
HCM (5% truck, $f_{hv}=0.975$, $f_p=0.98$)	2,210	2,210	HCM, 2010
HCM (5% truck, $f_{hv}=0.975$, $f_p=0.95$)	2,140	2,140	HCM, 2010
Rakha	1,730	1,700	Rakha & Arafeh, 2010
Rakha (Removed spillbacks)	1,800	1,725	Rakha & Arafeh, 2010
SERPM coded	2,142	2,142	Cambridge Systematics, 2008
Breakdown flow (15 minutes average before breakdown happens)	1,840	1,810	Elefteriadou and Lertworawanich, 2003
Queue discharge	1,625	1,630	Elefteriadou and Lertworawanich, 2003
Queue discharge (Removed spillbacks)	1,710	1,680	Elefteriadou and Lertworawanich, 2003
Maximum 5 minute interval observed (averaged over selected days)	1,930	1,925	Dervisoglu, 2009
Maximum 15 minute interval observed (averaged over selected days)	1,845	1,820	Dervisoglu, 2009
Maximum hourly averaged over selected days	1,745	1,745	Chao et al, 2005
Top 1% of hourly volume over all selected days	1,775	1,880	Jia et al., 2010
Volume associated with maximum occupancy in fundamental diagram	1,825	1,810	Van Arem & Van Der Vlist, 1992

Figure 5-3 illustrates how the Rakha model fits the observed data. The parameters that can be estimated based on this model are the capacity (the apex of the fitted model), jam density, free-flow speed, and speed at capacity.

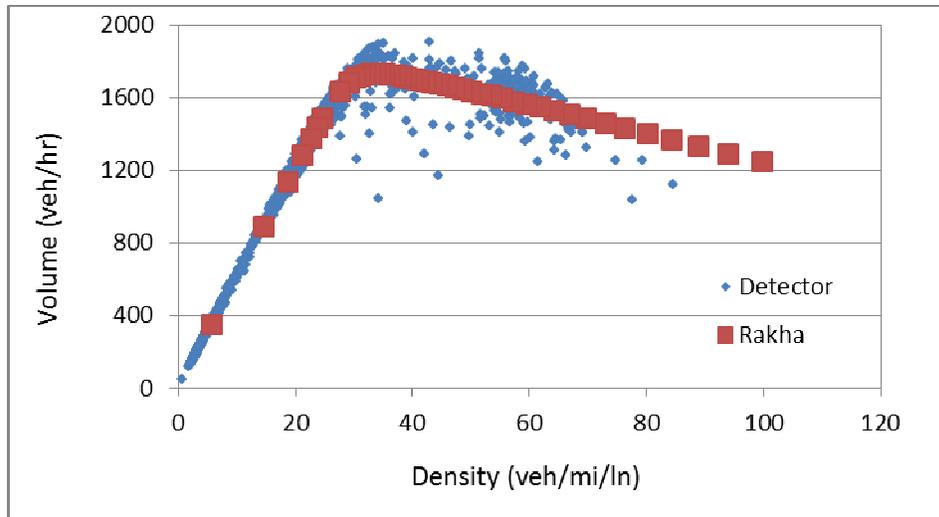


Figure 5-3 Rakha Model Fitting for Capacity and TFM Parameters Estimation

Figure 5-4 shows how the pre-breakdown flow and queue discharge rates were identified. In this study, the average of flow rates in three intervals before the speed drops due to breakdown is considered as the pre-breakdown flow.

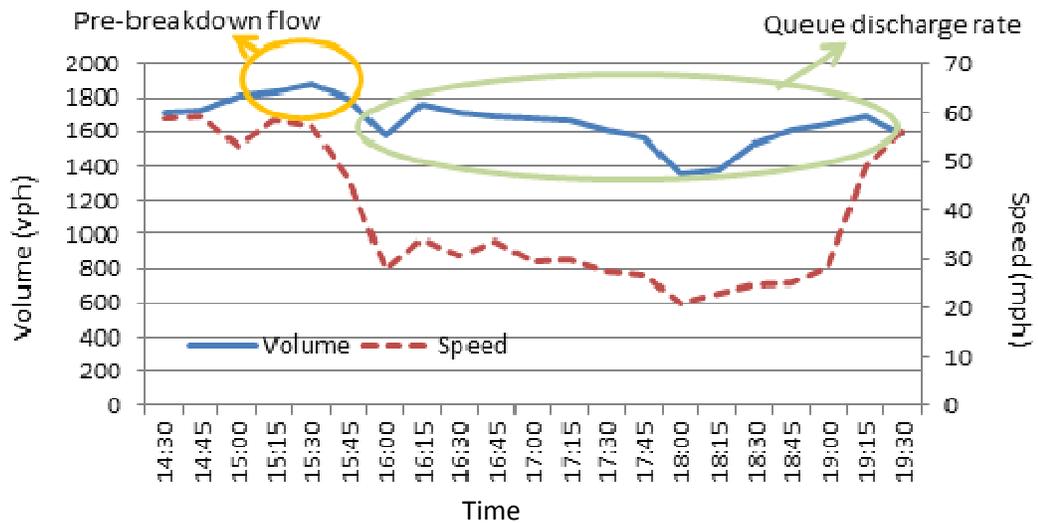


Figure 5-4 Demonstrating Breakdown Flow and Queue Discharge Rate

The capacity values discussed above are for the general purpose lanes of the corridor's cross-section. As stated earlier, I-95 also includes managed lanes that are separated from the general purpose lanes by soft barriers. Since congestion is avoided in the managed lane (ML) by toll value, there are not enough observations to estimate capacity from the real world. Based on literature, 99.5% of observed volume can be used as capacity. In this study, this value is almost 1,700 vph. Washburn et al. (2010) mentions capacity values ranging from 1,600 vph to 2,100 vph for existing managed lane facilities across the country.

5.4. Coded Capacity Impacts

The purpose of the discussion in this section is to illustrate the importance of coding capacity values estimates based on field measurements as input into dynamic traffic assignment (DTA) tools, particularly when there is evidence that the modeled corridor capacity is lower than the HCM-based estimates. It also demonstrates the shortcomings of utilizing static assignment for assessing managed lane utilization, even when the correct capacity values are coded, and subsequently illustrates the need to utilize DTA modeling for such assessments.

To illustrate the difference in the performance of different traffic modeling approaches, the volumes on the general purpose lanes and managed lanes were forced, in all modeling approaches, to resemble as much as possible real-world measurements based on detector data. For these fixed volumes, this study compared the travel times estimated based on the traffic flow models in static assignment with HCM-based capacity, static assignment with measured capacity, DTA with HCM-based capacity, and

DTA with measured capacity. Figure 5-5 shows the speed contour maps of the modeling results. This figure clearly shows that the only model that was able to replicate the real-world bottlenecks at Stations 12 and 20 was the DTA with measured capacity.

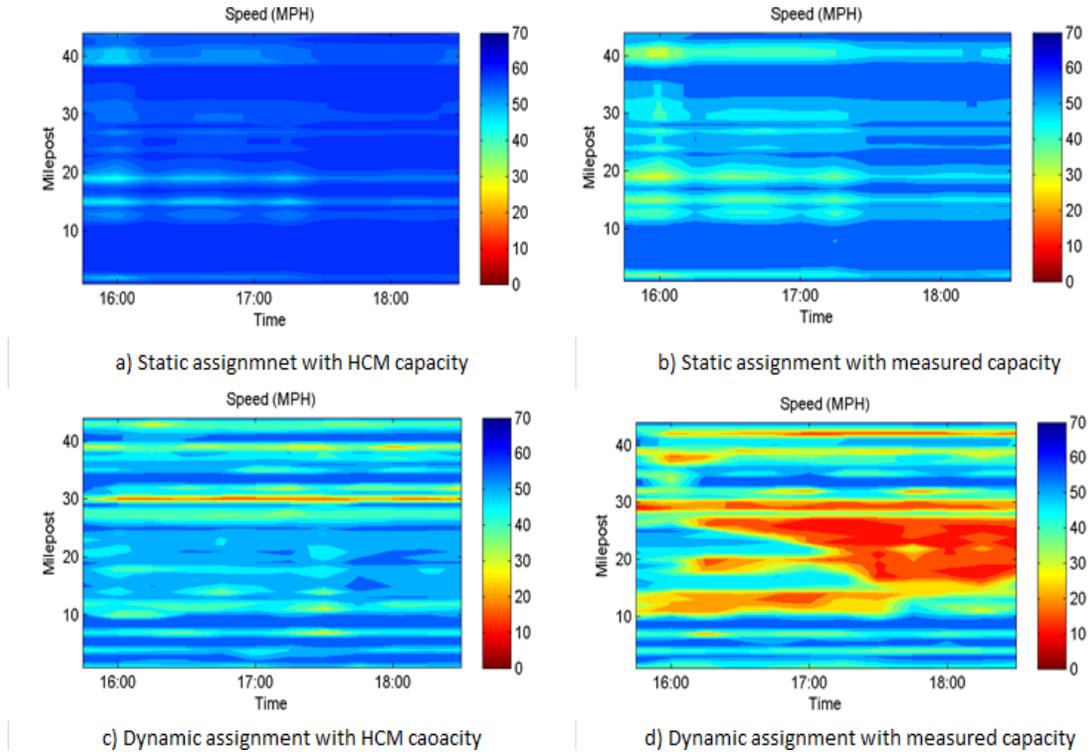


Figure 5-5 Speed Contour Maps for Static and Dynamic Traffic Assignment with Different Capacity Values

Figure 5-6 shows the difference in travel time between general purpose lanes and the managed lanes for the four modeling approaches. This figure confirms that the only model that could show the congestion observed in real-world conditions is the DTA model with the measured capacities. In static assignment, no queuing is assessed and the travel time is calculated based on a simple BPR curve. The change in the value of the capacity in static assignment does not have a significant effect on the modeling results. It

is also important to point out that in the DTA tool, when using the regional network capacity, no queue is formed; therefore, the results are similar to the static assignment tool.

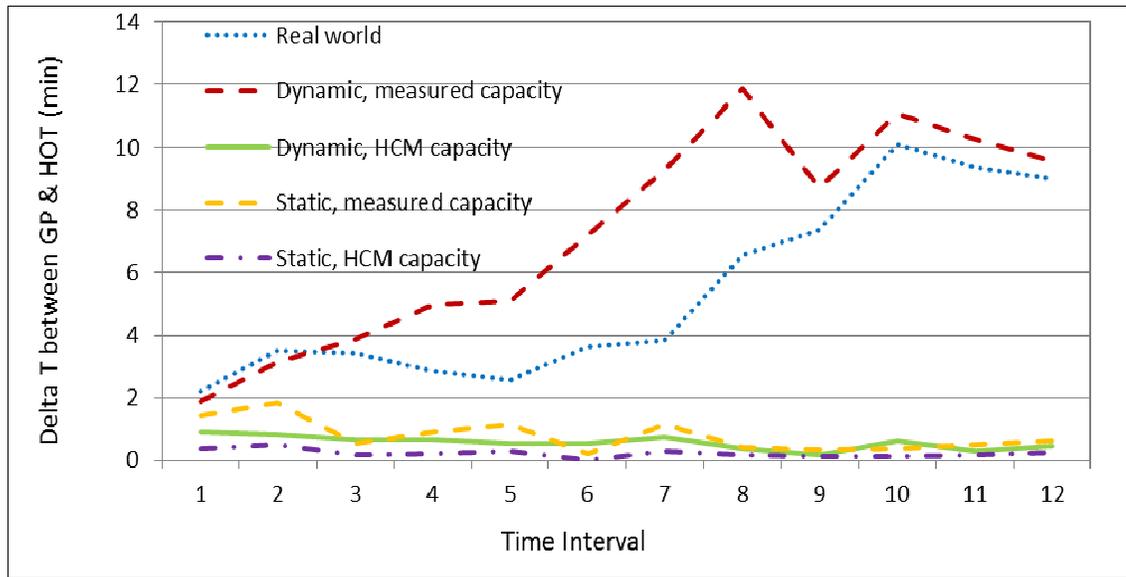


Figure 5-6 Travel Time Difference Between GPL and ML

The findings above are important because the difference in travel time between general purpose and managed lanes is used in the modeling process to assess the proportions of traffic that utilize the managed lanes, either based on user equilibrium assignment, a willingness-to-pay table, or a logit model combined with the assignment. This importance is further illustrated by feeding the difference in travel time results from Figure 5-6 to a willingness-to-pay table derived in a previous study (Ruegg and Puppala, 2013), so as to determine the change in the estimated percentages of traffic willing to use the managed lane. Assuming a \$1 toll for this segment, the percentage of drivers who are willing to pay the toll is calculated based on the willingness-to-pay curve. This calculation is based on toll value (in cents) divided by the saved travel time (difference

between general purpose and managed lane travel times). As is shown in Figure 5-7, the only model that was able to produce the expected results is the DTA model with the measured capacity.

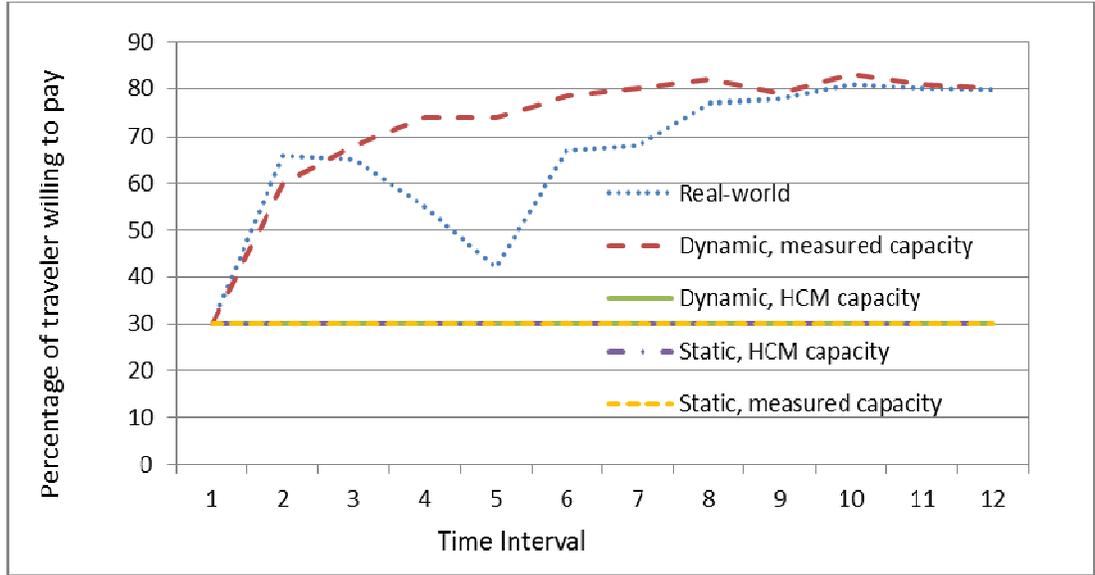


Figure 5-7 Percentage of Travelers Diverting to ML

It is worth noticing again that one of the congestion spots in this network is caused by a spillback from an off-ramp that causes low speeds in the two left lanes (the I-95 Northbound off-ramp to the Turnpike). Since the utilized DTA tool (Cube Avenue) does not support lane-by-lane modeling, it is not possible to correctly replicate that location, because the queue in the model first fills up the whole segment (including 5 lanes) before backing up to the upstream link. In the real-world, only the two left lanes are blocked. If replicating the congestion at such locations is important to a study, a tool that better handle this situation or multi-resolution analysis should be considered.

5.5. Other Traffic Flow Model Parameter Estimation

The Cube static assignment utilizes the widely used BPR traffic flow model to estimate the travel times during assignment for the whole analysis period, normally a peak period in case of time-of-day demand forecasting. On the other hand, Cube Avenue utilizes a mesoscopic simulation model to estimate the system performance at short time intervals during the simulation. The model generates individual vehicles and models and their interactions based on a TFM, with the performance further assessed using queuing analysis. Although the default traffic flow model is the BPR, the Cube script provides the flexibility to implement any desirable TFM. It should be emphasized, however, that in Cube Avenue, the TFM only affects travel time calculation when demand is below capacity. After queue formation, the delay values are calculated based on queuing analyses and can only be affected by adjusting the link capacity and storage parameters by the user. In other words, travel time is divided to two parts of moving on the link, and waiting at the link entrance gate due to capacity or storage restrictions. TFM affects the moving time, but the waiting delay is calculated internally.

Figure 5-8 shows the effect of implementing different TFMs on travel speed at the bottleneck location. Akcelik, Van Aerde, Greenshields and BPR curves are compared in the figure. It can be seen that during congestion the travel time for all TFMs is almost the same, but before and after breakdown, the travel time is slightly different, with BPR producing the lowest value.

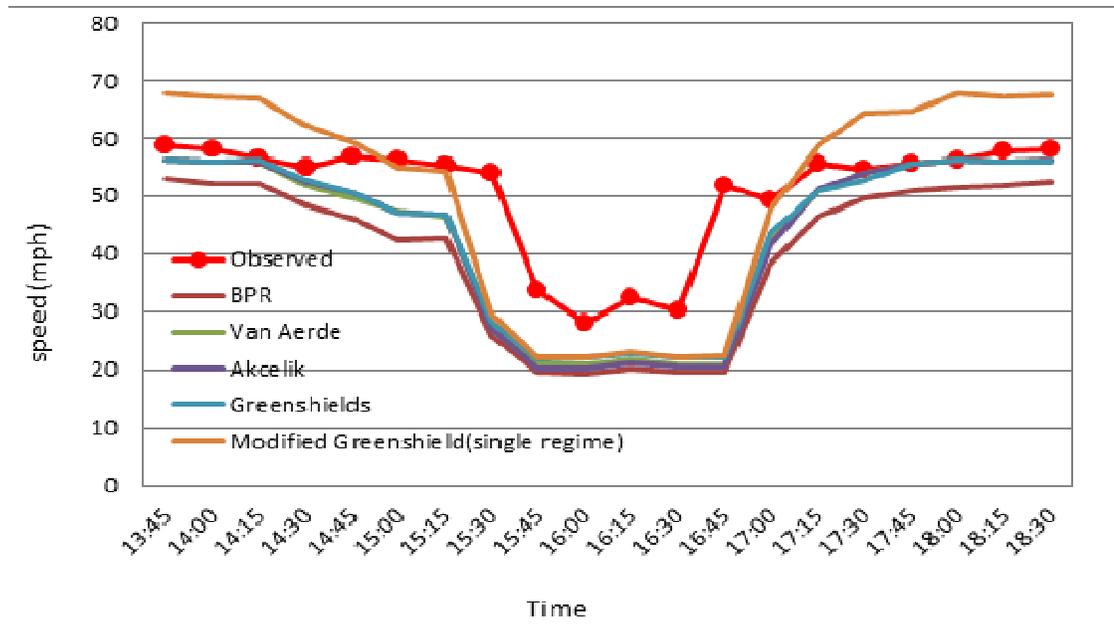


Figure 5-8 Effect of Implementing Different TFMs on Travel Speed

In Cube Avenue, it is documented that storage, along with capacity, are two constraints that limit the number of vehicles entering a link. The default value used in Cube Avenue is 190 veh/h/ln. This value is in the range of jam density rather than queuing density. Jam density is different from queuing density. Jam density is the density when all vehicles are stopped, while queuing density is the density of a moving queue. Coding the storage as jam density produces congestion spots with very low speeds (2 to 3 mph). ITS data however, shows a higher minimum speed in congested areas. In other words, cars move within queue, with a speed of 12 to 15 mph. This suggests that the storage should not be considered as jam density (completely stopped vehicles in a very congested network), but the queuing density should be used. This density can be calculated by dividing the volume by speed at the congested segment. This value is almost 3 times smaller than the jam density. By applying this value, the minimum speed increases and more closely resembles the observed speed. Queue length also more closely

resembles the real world. Figure 5-9 demonstrates the effect of jam density versus queuing density in replicating speed contour.

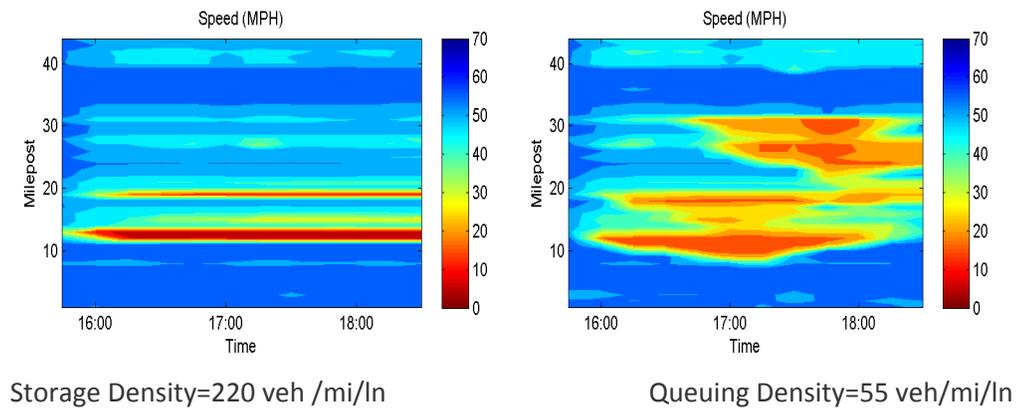


Figure 5-9 Difference Between Storage and Queuing Density on Travel Speed

The quality of the supply calibration is evaluated based on performance measures. Primary performance measures that evaluate how well the network replicates a real-world situation are link volume versus observed counts, and link speed versus measured speed. Several goodness-of-fit tests were suggested to measure the distance between simulated and observed volume. **Table 5-2** represents the most common goodness-of-fit measures that are used to assess network calibration.

Table 5-2 Goodness-of-Fit Measures

Goodness of Fit Measures	Formula
Root Mean Square Error (RMSE)	$\sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{N}}$
Root Mean Square Normalized (RMSN)	$\sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i y_i^2}}$
Percent Root Mean Square Error (% RMSE)	$\sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{N} * \frac{100 * N}{\sum \hat{y}_i}}$
Mean Absolute Error (MAE)	$\frac{\sum_i^n y_i - \hat{y}_i }{N}$
Scale	$\sqrt{\frac{\sum y_i^2}{\sum \hat{y}_i^2}}$
GEH (link-based)	$\sqrt{\frac{2(y - \hat{y})^2}{y + \hat{y}}}$

In formulas above, y is the simulated/estimated volume, and \hat{y} is the observed volume. Except for GEH, the above goodness-of-fit measures can be used to calculate the distance between estimated and observed values of other traffic measures such as speed, density, and queue length. GEH is an empirical-driven formula that has been proven useful for a variety of traffic analysis purposes, mainly for volume comparison purposes. A GEH of less than 5.0 is considered a good match between the modeled and observed hourly volumes (flows of longer or shorter durations should be converted to hourly equivalents to use these thresholds). According to the Federal Highway Administration

(FHWA), 85% of the volumes in a traffic model should have a GEH less than 5.0 (FHWA, 2007). GEHs in the range of 5.0 to 10.0 may warrant investigation. If the GEH is greater than 10.0, there is a high probability that there is a problem with either the travel demand model or the data. It should be noted though, that GEH may be misleading when used in assessing the accuracy of link volumes that are not in the same range. Consider a 3-lane segment, with observed 7,200 vph, and a 1-lane arterial with observed 800 vph. Assume that the simulated value for abovementioned segments are 5,400 vph and 600, respectively. The GEH value can be calculated as 22.7 for the first segment and 7.5 for the second case. This large difference cannot be justified given that in both cases, the simulated value is 75 percent of the observed value

In congested networks, volume replication must be coupled with speed or density replication. Considering only volume as calibration assessment criteria in such conditions can lead to a network that does not reflect the congestion patterns in the real world. It should be mentioned that detectors can only measure the served volumes, not the actual demands. Once demand exceeds capacity, the served volume starts decreasing with an increasing level of congestion and increasing density. This phenomenon cannot be captured by solely considering the detector volume. Congestion patterns in the model should reflect real-world conditions, spatially and temporally. Speed contour is a strong visual inspection tool for comparing congestion patterns of modeled and observed situations.

It is important that speed-volume profiles (temporal speed and volume for each link) at bottleneck locations should also be replicated. A comparison between modeled and observed speed/volume profiles (similar to Figure 5-4) can be used to evaluate how

well the model can replicate the following: starting and ending time of breakdown, speed and volume before breakdown, duration of breakdown, average volume and speed during breakdown, and covered speed and volume when the breakdown period is over.

Overall traffic measures such as VMT, VHT, and VMT/VHT can also be used for general evaluation of the calibration. It should be noted that abovementioned measures should be assessed in a calibrated network with fixed demand (calibrated demand).

5.6. Summary

Supply or network calibration in Cube assignment tools entails estimating capacity, free-flow, and traffic flow model parameters for each link in the network. A systematic, multilevel approach to network calibration is recommended in this study, with an increasing calibration scope in each level. The process starts at the level of separated bottlenecks, where the capacity is estimated by various methods based on field data. The network is gradually extended to connected bottlenecks, and then to the whole corridor and subarea coverage. The advantage of this approach is twofold: First, critical spots of the network can be better identified, analyzed, and replicated; second, a more reliable demand can be estimated for the smaller networks that are the focus of this study, which is very important in the iterative process of demand-network calibration. Focusing on isolated bottleneck locations and the freeway corridor for managed lane assessment enables the capturing of the interactions between supply and demand in addressing the causes for congestion. This is not feasible in more complicated networks.

The supply calibration performed in this study illustrates the importance of coding capacity based on detector measurements in DTA tools, particularly when there is

evidence that the modeled corridor capacity is lower than the HCM-based estimates. In the case explored in this study, it was found that the free-flow speed and, more importantly, the capacity were overestimated by the HCM procedures, resulting in incorrect travel times and congestion when used in the DTA model.

6. ORIGIN-DESTINATION MATRIX ESTIMATION

Dynamic traffic assignment requires trip matrices specified for short time intervals (e.g., 15 minutes or 30 minutes). These matrices are sometimes referred to as time-variant or dynamic trip tables. The derivation of these matrices is one of the most challenging aspects of dynamic traffic assignments.

The initial source of demand in this research is a trip (OD) table for a peak period, extracted from a regional demand forecasting model. The demand forecasting modeling process is a mature and well established process that produces behaviorally consistent results among different demand forecasting steps, including trip generation, mode choice, trip distribution, and trip assignment steps. These models are well calibrated based on real-world data and surveys. Therefore, they constitute a rich source of OD information with inherent consistency among trip generation, distribution, and assignment. These trip tables should be considered as an important source of demands. However, the trip tables need to be updated for operational purposes due to the necessity for shorter time intervals demand and the need for more focused validation of the demand for the subarea under consideration. The demand calibration or estimation step in this study aims to estimate the OD table for short intervals (15-minute intervals in this case) based on an initial matrix obtained from the demand forecasting model. The resulting matrices, when loaded onto the calibrated network, are able to replicate the observed link volume and congestion pattern.

As mentioned in Chapter 2, the demand estimation procedure can be significantly affected by the utilized network parameters, as well as by route choice (assignment)

parameters. On the other hand, calibrating network and assignment parameters requires correct demands. Thus, an iterative approach is needed for estimating the demands and network parameters.

The first step is to extract an initial OD matrix for the whole peak period from the regional model (Three-hour PM peak period from the SERPM model). The subarea boundary can be specified using the Cube Polygon feature or a GIS tool. Cube Voyager can then be used to extract the subarea network from the statewide model network using this predefined subarea boundary. The result of this extraction is a subarea network with new node and zone numbers and new trip table associated with this network.

The next step is to distribute this three-hour matrix over 15-minute intervals. This distribution was conducted to be consistent with the variations in observed volumes at uncongested locations at the beginning of the corridor, where detector volumes can represent actual demands (and not the capacity-restrained served volumes). The availability of these initial 15-minute interval matrices (referred to as factorized matrices in this study) allowed to start an initial network calibration, as described in Chapter 5, based on the 15-minute volume and speed data.

The next step is to adjust these matrices using the Cube Analyst static matrix estimation program. This matrix estimation process performs the estimation by considering a number of input parameters based on the static assignment of Cube Voyager. This process applies a maximum likelihood approach to estimate the trip tables that when assigned to the network paths, produce results that are close to the input data such as screenline counts, initial (seed) trip table, available route information, and zonal trip generation and attraction. This procedure is framed as bi-level optimization as

described in Chapter 2. The upper level formulation aims at estimating new trip tables based on maximum likelihood to minimize the difference between model output and the real-world or initial measures. In the lower level, the matrix estimated in the upper level is loaded on the network by an assignment module (Highway in case of static OD estimation) to produce link proportion matrix. This matrix includes the proportion of each OD that contributes to a link's volume. This matrix is then fed back to the upper level optimization for new matrix estimation iteration. It should be noted that if the initial OD pair is zero, it can never be raised automatically by the estimation process, unless the initial matrix is manually adjusted to reflect a non-zero value. The reason is that when the matrix with zero OD trip is loaded on the network, its contribution to link volume, and the resulting proportion matrix for that specific OD remains zero. Adjustment of OD pairs is essential and unavoidable therefore if there is evidence that their value is non-zero. This is just one example of the need for adjustments of the OD matrices, in addition to the implemented OD estimation model.

Since the STA runs over a single model period, each 15-minute interval must be run separately to estimate the OD matrix for the associated interval. The most important issue with STA in this process is its inability to capture queue spillback and make the resulting connection between consecutive intervals. This problem in the current study is minimized utilizing heuristics to account for queue presence.

The best approach to overcome STA limitations is to use the DTA instead of the STA as part of the least-square optimization to better account for traffic dynamics and travelers' behaviors. Thus, the intended next step was to use the Cube Analyst Drive procedure, which includes an OD estimation procedure that derives the time-variant trip

matrices based on minimizing the differences between the measured volumes and the volumes produced by the DTA, with consideration of initial trip tables resulting from the Cube Analyst estimation based on the STA. However, limitations were identified with the existing tool developed for this purpose, and modifications are proposed to improve the performance of this approach.

During the matrix estimation process, several manual adjustments and iterations were required. As demand changes, the network calibration may need to be slightly changed. The route choice behavior may also need to be adjusted, as is described in Chapter 7, as better OD estimates are obtained. Adjustments and fine-tunings are also needed to avoid unrealistic deviation from the initial matrix derived from the SERPM matrix estimation. These adjustments are iteratively and continuously performed during the matrix estimation process.

6.1. Static OD Estimation

The factorized 15-minute matrices derived based on 15-minute traffic counts are used in some studies as input into DTA models. However, these matrices can be further refined by utilizing a matrix estimation procedure based on traffic counts. Such a procedure would consider the deviations of the link volumes assigned by the model from traffic count measurements.

The static OD matrix estimation process is implemented using the Cube Analyst program, which is provided as an optional tool within the Cube modeling environment. Cube Analyst is a tool that estimates trip matrices based on the maximum likelihood technique, coupled with an optimization procedure. The tool utilizes data from different

sources and considers the different levels of confidence or reliability inputted by the user for these different sources. Not only can the data include traffic counts and prior (seed) matrices, but also partially observed matrices, zonal trip end (generation and attraction rate), vehicle routing, travel cost matrices, and even previously calibrated trip cost distribution functions.

Different sequences of processes for OD estimation were investigated in this study to determine how they impact the model's ability to replicate different measures of real-world traffic conditions, required memory and time, and deviation of the estimated OD from different sources of data. It was found that the best practice is to start with a factorized matrix, calibrating the network (supply), followed by static OD estimation, fine-tuning the network calibration, and then fine-tuning the ODs by performing dynamic OD estimation. Static matrix estimation was found to be the most essential step that could not be skipped. Running the dynamic matrix estimation (matrix estimation based on DTA) directly after the factorization step did not produce good results, possibly due to the immaturity of the dynamic OD estimation procedure in Cube Analyst.

Since Cube Analyst is based on static assignment, it deals with only one matrix at a time. Thus, it had to be run twelve times to obtain twelve 15-minute matrices in the three-hour period. Cube Analyst performs a set of iterative calculations that will automatically determine the statistically, most likely matrix for the set of input data values provided. The input data to Analyst can include the following:

- Screenline counts: These are observed link traffic counts at screenline locations. In cases that multiple user class matrices are estimated, the aggregated link counts should be split accordingly (i.e., each matrix class

should be associated with a class of observed counts). Each screenline can also be associated with a confidence factor. This feature enables the user to define the links that are more important to be replicated, or are associated with more reliable traffic counts. In this study, traffic counts for each 15-minute interval were obtained from ITS and PTMS detectors.

- Initial trip tables: One trip table is required for every user class. Each matrix can be associated with a confidence matrix, which contains different confidence level values for each OD pair. In this study, the initial 15-minute trip tables were obtained from the factorization process described earlier.
- Zonal trip ends: These are the total number of trips originating and terminating in each zone. Each zone can be associated with a confidence factor, based on the level of reliability or importance of preserving the total number of trips.
- Partial trip table: This optional input enables the user to incorporate any partial OD trips that are available from other sources such as Bluetooth readers, Electronic Toll Collection System, or OD surveys.
- Routing information: This information is provided by the assignment module. This input contains information of ODs that passed each link.
- Optimization parameters: These are parameters provided to set convergence criteria for optimization, and to set a weight that shows users' relative confidence on initial matrix versus screenlines. Higher weight shows that the user prefers not to deviate significantly from the initial matrix, even if the screenlines cannot be completely replicated. The appropriate confidence

values can be identified as part of the iterative process of the supply/demand calibration. These parameters are in a “control file” input to Analyst as a text file and the required and optional parameters can be easily edited by the user.

In this study, four groups of matrices are available from the regional demand forecasting model: Drive Alone (DA), Shared Ride of 2 occupants (SRP2), Shared Ride of three or more occupants (SRP3), and Truck. The I-95 ML policy does not differentiate between DA and SRP2 (e.g., both groups should pay the same toll to access ML). Therefore, the DA and SRP2 matrices are grouped together and are referred to as Single Occupancy Vehicle (SOV) in the assignment module.

As is described in Chapter 7, when using the willingness-to-pay approach to modeling ML, travelers are divided into two groups: toll payers and non-toll payers, based on the ratio of toll cost divided by saved travel time. It is assumed that SRP3 can use the ML without any cost or restriction, and trucks are not allowed to use ML. The summation of DA and SRP2 is split, based on the willingness-to-pay curve, into two groups: SOV_wo_Toll (non-toll payers) and SOV_w_Toll (toll payers). The SOV_wo_Toll and trucks are not allowed to use ML, but the other user classes choose between GPL and ML, based on the generalized cost function. The routing information is saved in binary “intercept” files associated with each user class. As a result, the aggregated link counts acquired from detectors were split accordingly into four user classes of SOV_w_Toll, SOV_wo_Toll, SRP3, and Trucks.

In the specific case of ML, which is the main interest of this study, the route choice behavior is highly complicated and has several parameters to estimate. Before running Analyst, the assignment process should be checked to confirm that it is able to

roughly estimate the portion of travelers that divert to the ML. If traffic assignment parameters, such as the willingness-to-pay curve, are not calibrated at this stage, the results negatively affect the OD estimation process. This creates another challenge, since a good assignment calibration requires a good demand estimation and vice versa. The network or supply calibration also affects the results. Thus, an iterative process is needed.

Another major consideration is capacity-constrained demands on congested corridors. Analyst is a robust optimization module that aims at replicating screenline volumes. However, in congested locations and periods; these volumes are the capacity-constrained served volumes. Thus, replicating these volumes based on counts will underestimate the demands. Northbound I-95 in the PM peak is a congested corridor, and as a result, the static OD estimation failed to produce the correct demands during the congested period. This problem can be solved by incorporating traffic measures that account for congestion presence, such as speed, density, or queue in the optimization tool. Due to the absence of these features in the current version, a method was developed to calculate the queue lengths, and it was added to the traffic counts in the screenline file.

The queue length on each link was estimated based on the level of congestion identified from the detector data. This value was added to the screenline volume count, and the static OD estimation was run again. The resulting demand was the input to the Cube Avenue module, and it was confirmed that it could better replicate real-world congestion patterns. Figure 6-1 demonstrates the effect of considering queue on congestion pattern replication. As can be seen, if OD estimation is solely based on screenline counts and no means of congestion is incorporated in the optimization, the estimated OD cannot produce the real-world traffic condition.

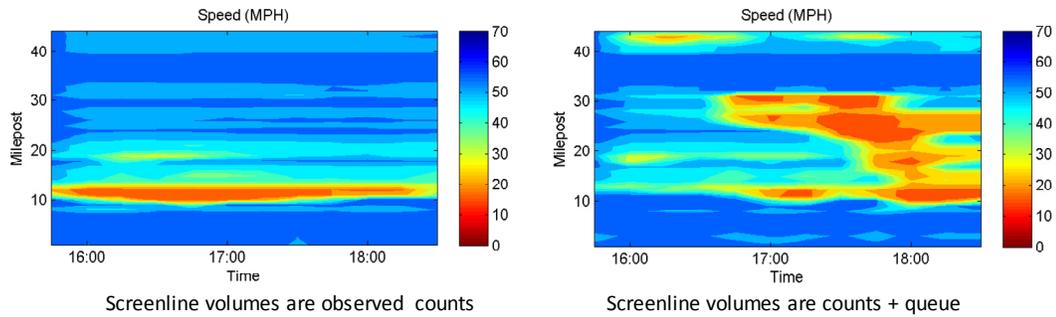


Figure 6-1. Effect of Incorporating Queue in OD Estimation

Less perfect replication of screenline volume is expected when the result of OD estimation (optimization) is manually adjusted. In this case study, manual adjustment increases the RMSE of screenline volume versus model output volume, from 56.8 to 61.8.

A new tool called Analyst Drive was recently developed by Citilabs. Analyst Drive can be used for estimating OD matrices based on static and dynamic assignment. There is a keyword in the control file as “OD TYPE”. Setting this value to zero runs static estimation, and setting a value of one runs dynamic assignment. In this study, Analyst and Analyst Drive were both run for static OD estimation with the same input and with the default parameters. Figure 6-2 displays the demand for a specific OD pair over 12 intervals. Figure 6-3 to Figure 6-5 present the real-world replication of mainline volumes when utilizing factorization, Analyst and Analyst Drive, respectively. For better tracking and visualization, flow rates from different screenlines are color coded in these figures.

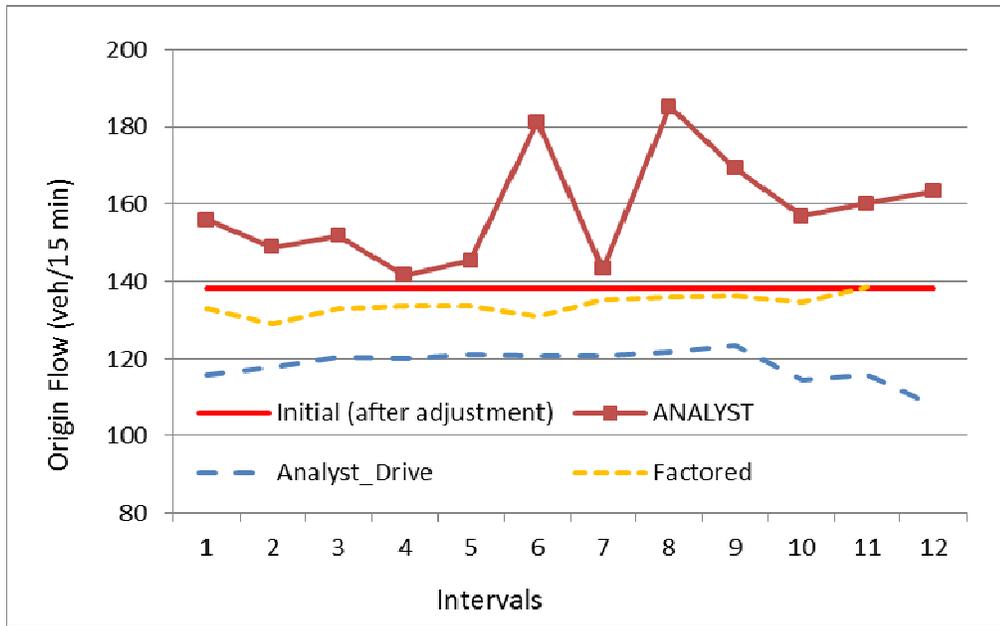


Figure 6-2 Temporal Profile of Initial, Analyst, and Analyst Drive OD

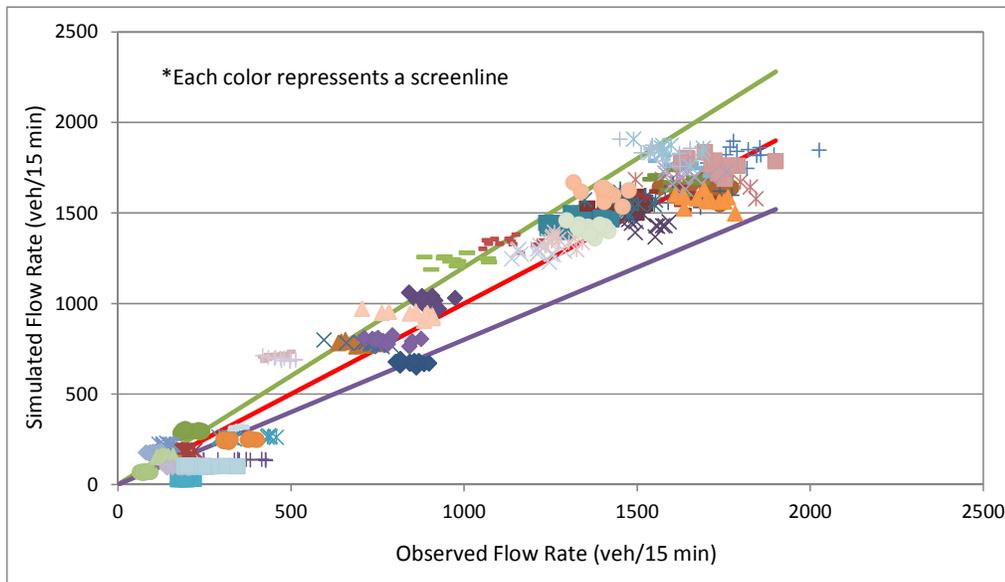


Figure 6-3 Screenline Volume Replication by Factored Regional Matrix

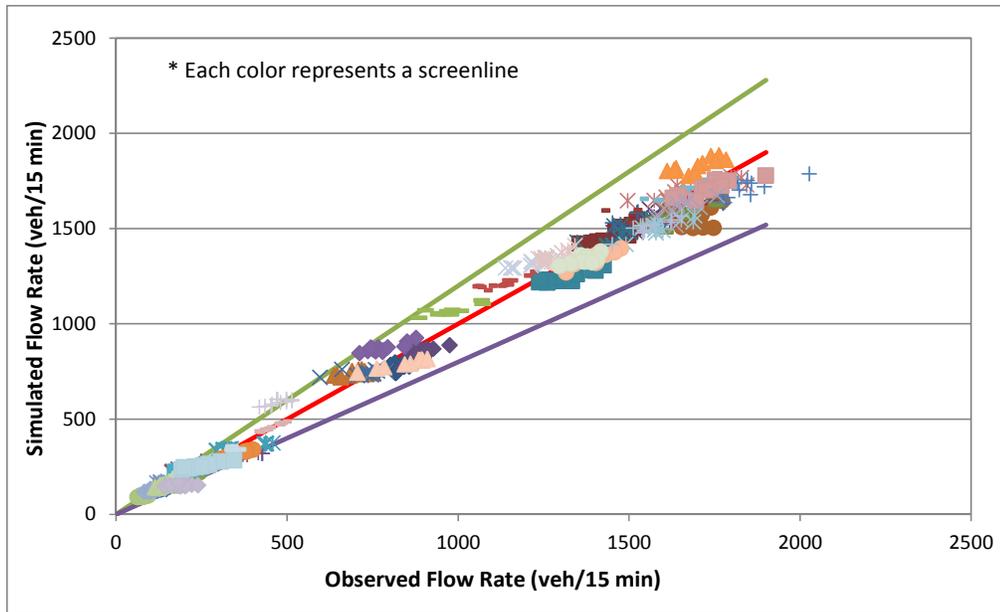


Figure 6-4 Screenline Volume Replication by Analyst

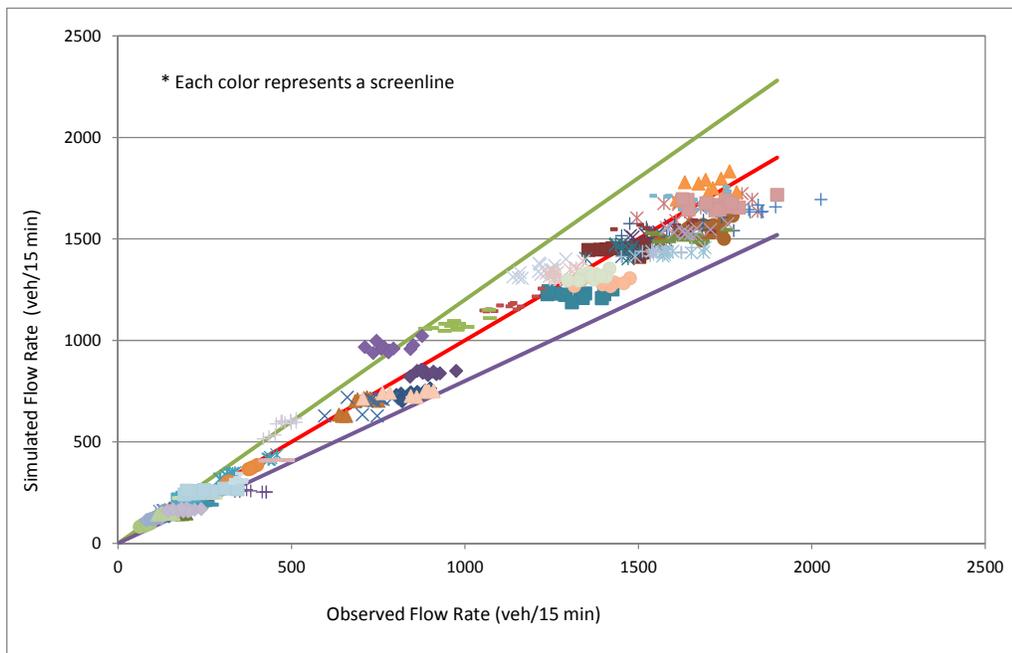


Figure 6-5 Screenline Volume Replication by Analyst Drive

Figure 6-6 compares the flows of one specific origin to all destinations, in the initial OD matrix and the estimated ones by factorization, Analyst and Analyst Drive.

This figure shows that Analyst tends to focus on replicating screenline counts, sometimes at the expense of deviating significantly from the initial matrix.

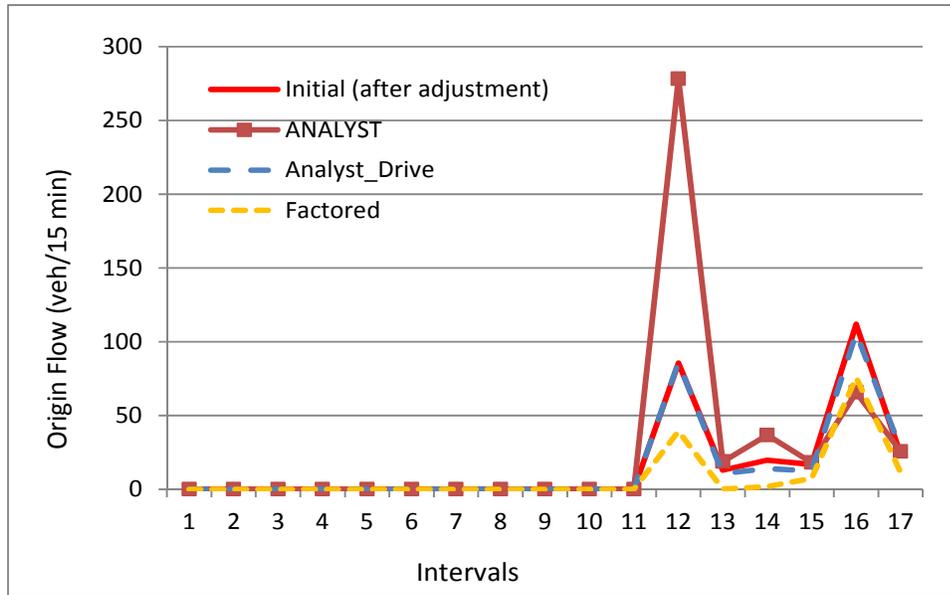


Figure 6-6 Comparing Initial and Estimated OD for One Pair

6.2. Dynamic Matrix Estimation

The 15-minute matrix estimation that uses Cube Analyst is expected to represent significantly better demand estimations than the ones produced from the factorized matrices that were used as inputs to the Cube Analyst (in the estimation process). However, the Cube Analyst process utilizes demands from the static assignment during the optimization process. The most important concern with using the STA in this process is its inability to capture queue spillback in space and time. In the current study, this problem could only be partially addressed by utilizing heuristics to account for queue presence. Thus, this study also investigates the use of the Analyst dynamic OD estimation process that utilizes routing information from Cube Avenue in the optimization process.

The dynamic OD estimation follows a very similar process, as described in the previous section. Instead of running Analyst with Highway assignment module for 12 consecutive periods of 15-minute intervals, Analyst Drive runs during the whole model period, coupled with the Cube Avenue assignment module. In the single Cube Avenue run, the model period is divided into 15-minute intervals. This procedure is supposed to be superior to static OD estimation, because Cube Avenue models the queues and queue spillbacks, and thus can capture the effects of congestion on subsequent time intervals. However, without incorporating density or speed, dynamic OD estimation may also underestimate the demands under congested conditions. The dynamic OD estimation module in Cube package is not as mature as the static matrix estimation module, and has the following limitations:

- Analyst Drive for dynamic OD estimation does not incorporate zonal trip ends. Zonal trip ends are usually available from ramp count data and provide valuable, reliable data about origin and destination trips.
- Partial trips cannot be incorporated into the current version of Analyst Drive for dynamic OD estimation.
- The confidence matrix associated with each input matrix cannot be incorporated into Analyst Drive for dynamic OD estimation.

Due to the aforementioned limitations and the results of running dynamic OD estimation in this study, it is suggested that the user should use caution when utilizing the dynamic OD estimation module in Cube Analyst.

6.3. Performance Measures and Matrix Adjustment

When calibrating simulation, demand, and assignment parameters, a distance function between simulation outputs and field measurements is minimized. This function can include different measures, such as link volumes, OD demands, link speeds and/or densities, etc. Limiting the function to replicating link volumes, as is the case in many studies, can be misleading and fail to produce the correct demands or congestion patterns. Most OD matrix estimation methods are based on link traffic volumes and initial OD matrices. Data on speeds, densities, queue lengths, OD routes, or zonal trip end rates should be incorporated into the calibration process to better replicate real-world traffic conditions. There are different ways to incorporate this information into the calibration process. They can be included in the objective function of the optimization or be a part of a manual adjustment or a heuristic procedure outside the optimization tool.

As mentioned in Chapter 2, demand estimation is an underspecified problem. This means that the number of equations (the link counts) is usually much lower than the number of unknowns (OD pairs). Hence, different OD estimates may produce the same link volumes. It is important, therefore, to manage the estimation process to ensure the reasonableness and the correctness of the estimated demands. Regarding the dependency between demand estimation and traffic assignment, a wrong estimation of OD pair demands can sequentially propagate during the calibration process. In order to limit the systematic errors in OD estimation, the consistency and reliability of the adjusted OD pairs should be checked against different sources of data, such as trip end rates or specific

route volume information. Following is a list of criteria that were identified in this study to justify manual adjustments of the estimated demands:

- Deviation from the initial matrix: It might be helpful to preserve certain structure or information that the initial matrix (subtracted from regional forecasting model) contains, such as the proportion between the total trips of the DA, SRP2, SRP3 and Truck user classes. Another example is to preserve the split between two major destinations in the network, such as I-95 and the Florida Turnpike northbound in the test network.
- Route information: There might be reliable information about specific route trips, which are necessary to replicate.
- Zonal trip end: On-ramp and off-ramp counts, in the absence of queues, can be reliable sources for origin and destination demand estimation, particularly in the case of linear corridor modeling. Thus, replicating these counts justifies the manual adjustment of the OD volumes.
- General temporal uniformity: There are no expectations of seeing unrealistically high rises or drops in the volumes of OD pairs in sequential intervals. In the Analyst optimization tool that is based on static assignment, the temporal variation cannot be controlled. The optimization process can achieve totally different local optimal solutions for sequential intervals, since the optimization does not guarantee achieving global optimal. To minimize the OD matrix variations between sequential intervals, the static OD estimations for different intervals were run with an identical initial matrix.

After several OD estimation trials and matrix adjustments, one matrix was selected as a good initial matrix, and was duplicated over twelve intervals.

Since the demand estimation is underspecified and may result in a local minimum, it may be helpful to force the optimization to start the search from a certain point, more specifically, to restrict some of OD pairs from varying during the optimization. Manually adjusted values should be inserted in the process again for a new run of the OD matrix estimation. Different approaches can be used to combine the estimated and adjusted values to control the deviation from the general structure of the initial matrix, such as Kalman filtering, Bayesian inference, and MSA. The adjusted and combined values will then be fed back into the estimation process. Modifications to the existing OD estimation process are recommended so as to allow the user to have the flexibility required to incorporate additional factors as limiting criteria in the objective function (based on the analyst's knowledge), to minimize the need for manual adjustments.

6.4. Recommendations for OD Estimation Improvement

Additional recommended improvements to the OD estimation process are listed as follows:

- Incorporating speed, density, and/or queue length in the objective function of the OD estimation process,
- Allowing the user to specify lower and upper bounds for each OD pair cell (there is already a global parameter that is applied to all cells yet cannot be varied by cells).

- Allowing the user to better control the temporal variability of the results.
- Allowing the user to keep the proportionality between specific OD pairs (e.g., from all of the trips originated from I-95, with 30% directed to SR 836 and 30% destined to the Florida Turnpike).
- Incorporating zonal trip end, partial matrix, partial trips, and confidence matrix for dynamic demand estimation.

6. 5. Summary

The process of converting regional OD matrices to OD matrices that can be used as inputs to DTA is described in this chapter. This process includes a combination of the factorization, static OD estimation, and dynamic OD estimation steps. Considering the issues with the current state of the available dynamic OD estimation tools, static OD estimation was considered the most essential step, in which ODs are estimated in a way that, when loaded on the network, produce real-world measures such as link volumes. Improved dynamic OD estimation process, however, are expected to produce better results than static OD estimation.

It was found that in congested networks without incorporating measures of congestion such as queue, speed, or density, the resulting demands are underestimated and cannot produce real-world traffic conditions. To enhance the model for congestion pattern replication, queue was estimated and incorporated into the process.

The estimated OD matrix should be checked to avoid any structural deviation from the regional matrix, unless there is strong evidence for a necessary change in specific OD pairs. A minimum amount of manual adjustment of the estimated OD table is

necessary and inevitable. Manual adjustment can be minimized by automatic methods such as successive average, or machine learning approaches. OD estimation results should be fed back to the other two steps of the calibration, including network calibration and route choice model calibration. Then, the demands can be further adjusted based on the results.

7. TRAFFIC ASSIGNMENT

Tests and comparisons of traffic assignment methods, as well as calibration results, are presented in this chapter, and they are based on the I-95 linear (corridor) network and trip tables that were calibrated as described in Chapters 5 and 6. A side-by-side comparison between STA and DTA is conducted with the STA running twelve times, once for each of the 15-minute trip tables during the PM peak period. Therefore, the output file contains volumes and speeds for twelve time intervals that are used in the comparison with real-world data and DTA results. It should be mentioned, however, that these runs are independent from each other, and the run for one interval is not affected by the results of the previous interval because STA is not capable of modeling these interactions between time intervals. Two different approaches for ML assignment are discussed and evaluated based on performance measures including replication of real-world volume and speed, ML demand, model convergence, and stability.

7.1. Route Choice Behavior Modeling

Two different approaches are investigated for modeling the route choice behavior. The first approach involves adding the toll cost to the generalized utility function of the link. In this approach, referred to as “Generalized Cost Function”, selecting the route is solely governed by the user equilibrium assignment procedure, based on the generalized costs of different paths. In the second approach referred to as the “Willingness-to-Pay Curve”, prior to the assignment, travelers are divided into two groups: a group that will

not choose to pay the toll and is limited to using GPL; The other group is eligible to use ML based on the willingness-to-pay curve, but the final decision to use either the ML or GPL depends on its origin and destination points (whether there are proper entry/exit points if they take the ML), as well as on the difference in the travel time between ML and GPL according to the user equilibrium process. Figure 7-1 demonstrates the assignment procedure based on willingness-to-pay curve.

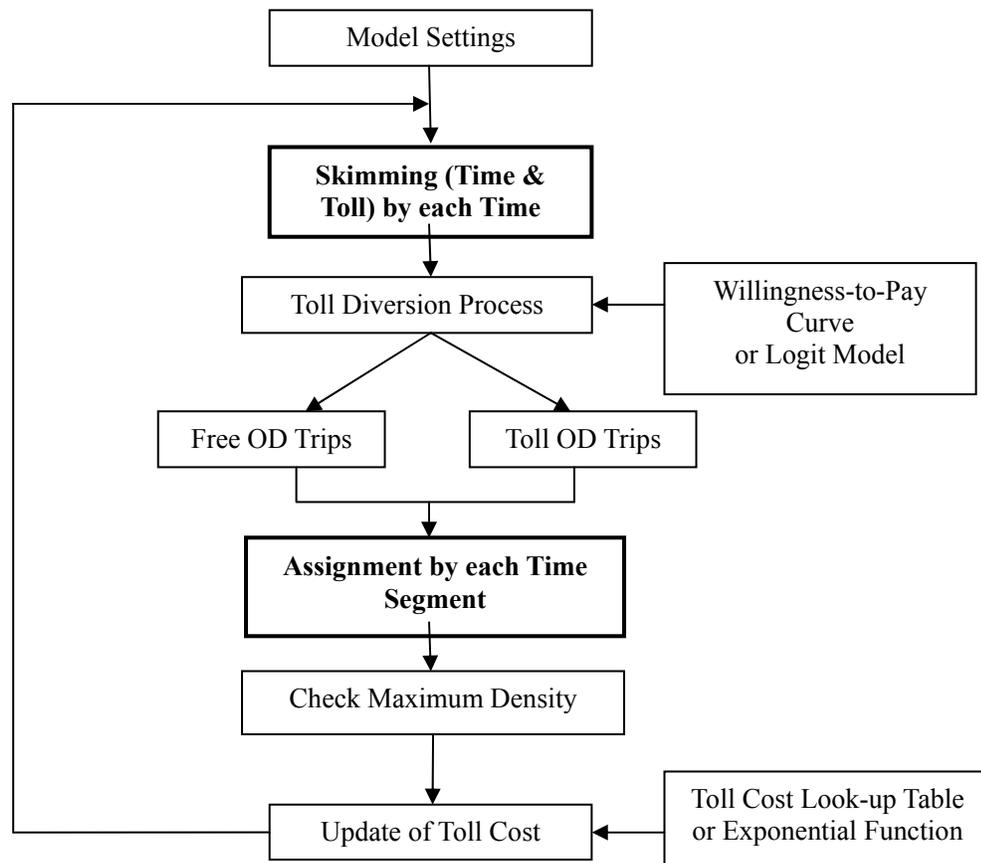


Figure 7-1 Managed Lane Model Based on Willingness-to-Pay Curve

After setting network parameters, the path-building process is performed to obtain the impedance values, such as the travel time (min) and cost (cents) for each origin-destination pair. These skimming values are used to compute the ratio of toll in

cents over the time saved between free and toll routes from origin to destination as follows:

Toll Cents per

(7-1)

Next, the toll trip share (%) can be obtained by looking up the willingness-to-pay table (see example in Table 7-1). For example, suppose that a driver can travel the free road in 25 minutes, while the driver also has an opportunity to use the toll road with a travel time of 20.78 minutes by paying \$1 as a toll cost. In this case, the toll cents per minute saved is 23.7 cents per minute ($=100 \text{ cents}/(25-20.78)\text{min}$). Thus, the user's probability of using the free road is 85%, based on the willingness-to-pay table, while the probability of using the toll road is estimated at 15% ($=100\%-85\%$).

Table 7-1 Initial Not-Willing-to-Pay Proportion for Cost per Time Saved by Demand Category

Toll Cent per Minute Saved	Demand Category					
	1	2	3	4	5	6
0.0	5.0	5.0	5.0	5.0	5.0	5.0
8.0	50.0	50.0	50.0	50.0	50.0	50.0
10.0	60.0	60.0	60.0	60.0	60.0	60.0
16.3	75.0	75.0	75.0	75.0	75.0	75.0
20.0	81.7	81.7	81.7	81.7	81.7	81.7
23.7	85.0	85.0	85.0	85.0	85.0	85.0
31.4	90.5	90.5	90.5	90.5	90.5	90.5
41.7	95.0	95.0	95.0	95.0	95.0	95.0
51.8	96.0	96.0	96.0	96.0	96.0	96.0
58.3	98.0	98.0	98.0	98.0	98.0	98.0
66.7	98.8	98.8	98.8	98.8	98.8	98.8

In performing assignment utilizing both of the abovementioned methods, the toll is updated for each interval, based on the maximum density of the ML so as to preserve the desired level of service in ML. The schedule of the value of toll based on density was calibrated based on available charged toll data from the FDOT District 6 traffic management center, as well as based on ITS volume and speed data. Other important calibration parameters of the assignment are the value of time and the shape of willingness-to-pay curve.

7.2. Derivations from Observed Data

Implemented toll data, coupled with microwave detector data, is used to calibrate the toll-density curve (table). Table 7-2 includes the default toll values, which is a simplified version of the table that FDOT District 6 TMC uses to calculate the toll and it does not completely replicate the current I-95 toll table.

Table 7-2 Default Toll Values Based on the ML Maximum Density

Level of Service	Road Density		Toll Cost (\$)	
	Minimum	Maximum	Minimum	Maximum
A	0	11	0.25	0.25
B	12	18	0.5	1.25
C	19	26	1.5	2.75
D	27	35	3	3.75
E	36	45	3.75	6
F	>45		6	7

It was found that this table overestimates the toll values.

Figure 7-2 demonstrates the difference between the real-world charged toll and the calculated toll values based on the default toll-density in Table 7-2 for May 11, 2010. The calculated toll was obtained by estimating the density as the volume over speed ratio.

The density was calculated at each ITS detector along the managed lane for each of the 15-minute modeling intervals. The maximum density value along the eight-mile length of the managed lane was then used to calculate the toll costs, based on Table 7-2.

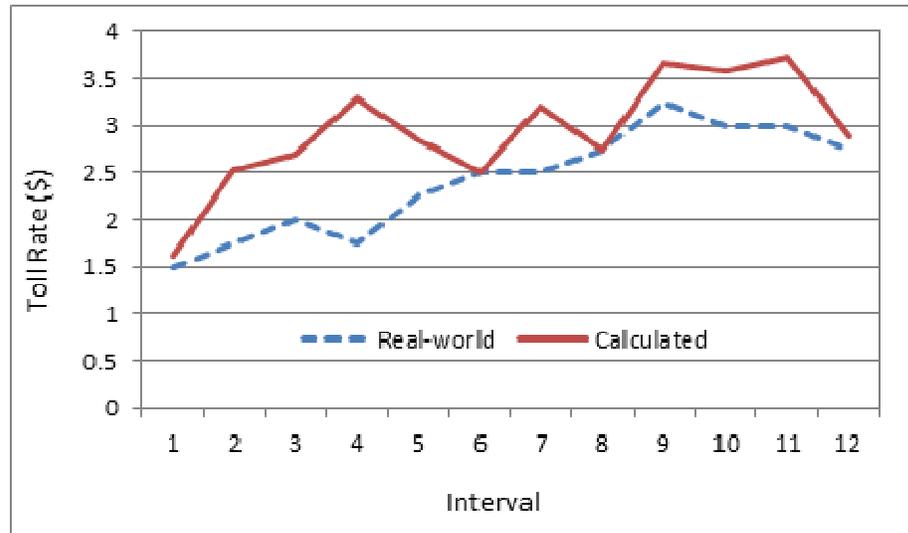


Figure 7-2 Comparison Between Implemented and Calculated Toll for a Lightly-Congested Day

Error! Reference source not found. and **Figure 7-3** demonstrate the same comparison for severely congested intervals (not just for one specific day). The values for the calculated tolls are derived from **Table 7-2**. It is clear that the utilized toll table overestimates the toll values.

Table7-3 Implemented Toll Value for I-95 Northbound

Day	Time	Toll Rate (\$)	Maximum Density from ITS Data (veh/mile/lane)	Calculated Toll (\$)
6/3/2010	3:41:00	2.50	31.46	3.45
6/3/2010	3:56:00	2.50	33.21	3.72
6/3/2010	4:11:00	3.25	37.01	4.31
6/3/2010	4:26:00	3.75	44.22	5.43
6/3/2010	4:41:00	4.75	60.84	7.00
6/3/2010	4:56:00	5.50	59.04	7.00
6/3/2010	5:11:00	5.00	66.10	7.00
6/3/2010	5:26:00	4.50	79.53	7.00
6/3/2010	5:41:00	5.00	77.43	7.00
6/3/2010	5:56:00	4.50	68.13	7.00
6/3/2010	6:11:00	4.50	67.29	7.00
6/3/2010	6:26:00	3.75	60.64	7.00
6/3/2010	6:41:00	3.25	51.99	6.63
6/8/2010	3:41:00	1.75	33.51	3.77
6/8/2010	3:56:00	1.75	31.91	3.52
6/8/2010	4:11:00	2.00	51.65	6.58

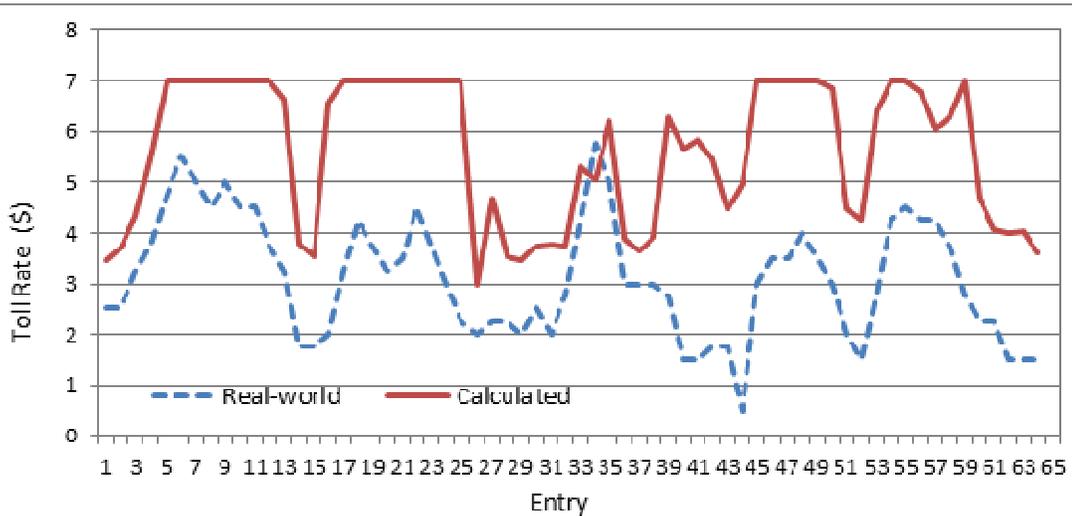


Figure 7-3 Comparison Between Implemented and Calculated Toll for Highly-Congested Intervals

Based on the abovementioned data, a new toll-density relationship was developed as shown in Figure 7-4. It should be noted that the toll schedule in real world, is based on a more complex lookup table that is difficult to implement. Also, the calibrated density value utilized to update the toll values in TMC might be different from the ones calculated in this study based on ITS data. To avoid complications, a simplified toll-density curve is developed in this study based on calibrating to real-world data. Figure 7-5 shows that the toll-density curve developed based on the observed data better replicates the real-world diversion to the ML. It should be noted that the results presented in Figure 7-5 are for a model with calibrated network, demand and route choice parameters. Therefore, the difference between curves is merely due to the toll schedule and no other factor.

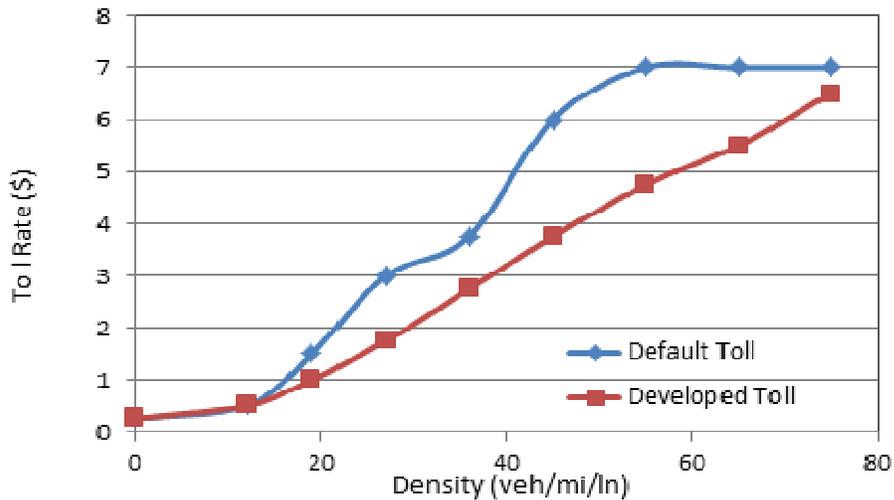


Figure 7-4 Default and Developed Toll Density Curve

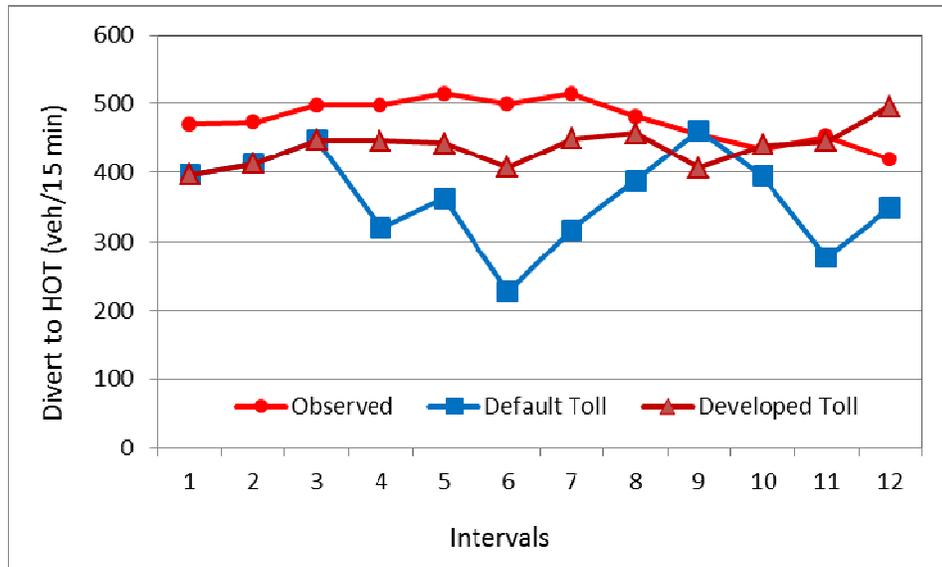


Figure 7-5 Comparison of Diverted Volume to ML for Different Toll Curves

7.3. Calibrating the Value of Time

This section describes the approach to calibrate the value of time parameter. This parameter converts monetary value of toll cost into equivalent time. This equivalent time can be then added to the utility function of the ML facility. If the summation of route travel time and the added equivalent time is still smaller than the congested time in GPL, ML is more attractive to the user. A value of time of \$30 per hour means that the user will pay \$30 to save one hour, or 50 cents for every minute of saved time. As discussed in Chapter 2, the value of time parameter might be interpreted as toll value that travelers will pay, not solely to save time, but to benefit the improved trip reliability, safety, and comfort that the ML facility offers. It was found that for the PM peak period, the average toll cost over several days in 2010 (excluding weekends) is between \$2 and \$3, with an average of \$2.30, as presented in Table 7-4.

Table 7-4 Implemented Toll Value for I-95 Northbound

Time (PM)	day 1	day 2	day 3	day 4	day 5	day 6	day 7	day 8	day 9	day 10	Average
3:26	1.00	1.50	1.50	1.50	1.75	1.75	1.75	1.75	2.00	1.75	1.63
3:41	2.00	2.50	2.00	2.25	1.50	1.75	2.00	2.75	2.50	2.25	2.15
4:11	1.50	3.25	1.75	1.75	1.75	2.00	2.00	2.00	2.00	2.00	2.00
4:26	1.50	3.25	1.75	1.75	2.00	2.00	2.25	2.00	2.00	2.00	2.05
4:56	1.50	3.00	1.75	1.75	2.00	2.00	2.25	2.25	2.25	2.00	2.07
5:11	1.50	3.00	1.75	1.75	3.00	2.00	2.25	3.25	2.50	2.00	2.30
5:26	1.50	3.00	2.00	2.50	3.00	2.25	2.25	2.75	2.75	2.25	2.42
5:41	1.75	3.50	2.00	3.00	3.50	3.25	2.5	2.75	3.00	3.00	2.82
5:56	1.75	2.50	1.50	3.00	3.75	3.00	3.00	3.25	3.50	3.00	2.82
6:11	1.75	3.75	1.50	3.50	3.50	2.75	2.75	3.25	3.00	2.75	2.85
6:26	1.50	3.75	1.50	2.75	2.75	2.00	2.75	3.00	2.50	2.00	2.45
6:41	1.50	3.00	2.25	2.25	2.25	1.50	2.25	2.25	2.50	2.00	2.17

The time saved by motorists based on real-world detector data for non-incident days is between 4 to 8 minutes, depending on the congestion level in the GPL for the day under consideration. Paying an average toll value of \$2.3 implies that travelers' value of time is \$17 to \$34. This value considers only the saved time and no other attractions of ML such as improved trip reliability and safety. The value of time in this study is estimated to be \$42, compared to the value of \$12.6 used in the SERPM model. This difference can be interpreted as the perceived benefits of using the ML beyond the absolute difference in travel time between ML and GPL.

Value of time of \$12.6, \$18.0, and \$31.0, \$42, and \$50 were used in sensitivity analysis. The results are displayed in Figure 7-6. From this figure it appears that values of time of \$42.00 produce good results. It should be noted again, that in absence of bias factors, this value accounts for factors rather than toll and saved time, such as travel time reliability, comfort, safety, and the travel time in past days, which include more congested days and incident days.

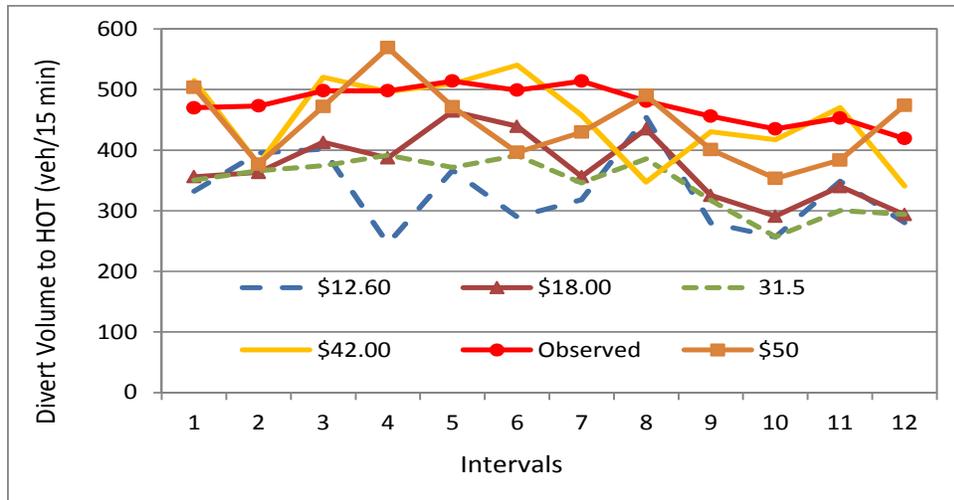


Figure 7-6 Comparison of Diverted Volume to ML for Different VOTs

7.4. Calibrating the Willingness-to-Pay Parameters

The willingness-to-pay curve defines the proportion of people that are *not* willing to divert to ML, based on the ratio of the toll value (in cents) divided by the saved time (in minutes), prior to user equilibrium-based assignment. Non-toll-payers based on willingness-to-pay curve, do not access to ML and are limited to use GPL only. Other travelers, however, chose their path based on their origin-destination (i.e., if there are proper entry/exit points if they select ML), and travel time on each path.

Figure 7-7 depicts different willingness-to-pay curves, and Figure 7-8 shows the diverted volume to ML. Based on the results in Figure 7-7 and Figure 7-8, Curve I was selected as the willingness-to-pay curve that best reflect real-world traffic count on ML. As can be seen, the shape of the calibrated willingness-to-pay curve is significantly different from the initial curve.

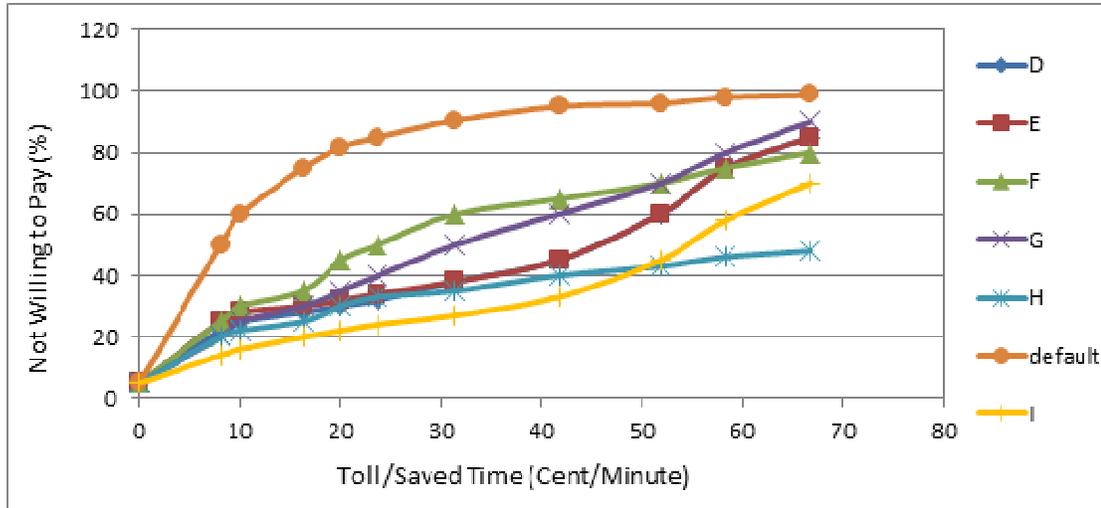


Figure 7-7 Different Shapes of Willingness-to-Pay Curve

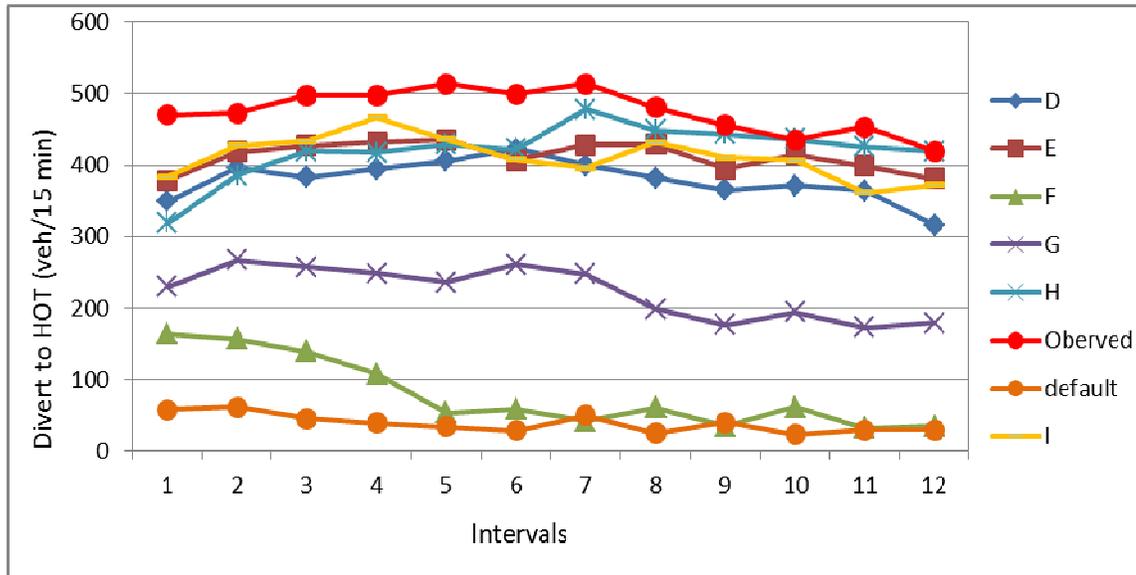


Figure 7-8 Diverted Volume to ML Associated with Willingness-to-Pay Curves

7.5. Comparing Static and Dynamic Traffic Assignment

This section demonstrates the difference between STA and DTA abilities to replicate the observed route choice behavior. Figure 7-9 and Figure 7-10 demonstrate

the difference between STA and DTA in predicting the divergence to the ML for the generalized cost function method and the willingness-to-pay curve method, respectively. The predicted divergence to the ML is also compared to the observed values derived from ITS data. As previously mentioned, the module in the Cube package for static assignment is called “Highway,” and the module for dynamic assignment is called “Avenue.” Highway module is run for twelve intervals, one for each 15-minute interval. Figure 7-99 and Figure 7-100 show that both approaches of ML modeling with Avenue produce results that are close to real-world measures. However, the Highway module is not able to replicate real-world measures. The main reason is that STA cannot model the variation in demand, queue formation and spillback, and the associated delays. The difference between GPL and ML travel time, and the resulting number of travelers that decide to choose the ML, is considerably underestimated by static assignment.

The figures also show that the generalized cost approach and the willingness-to-pay approach produce comparable results.

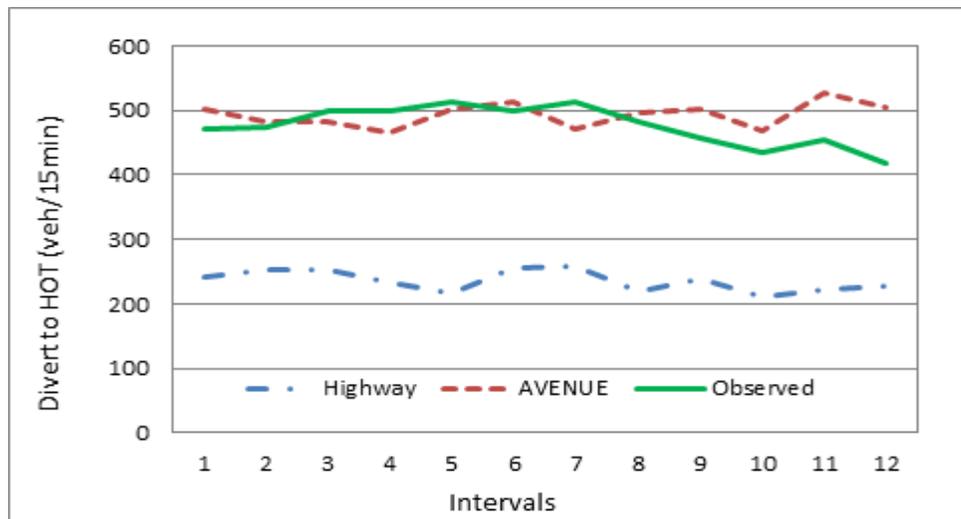


Figure 7-9 Comparison Between Modeled and Observed ML Volume for Generalized Cost Function Method

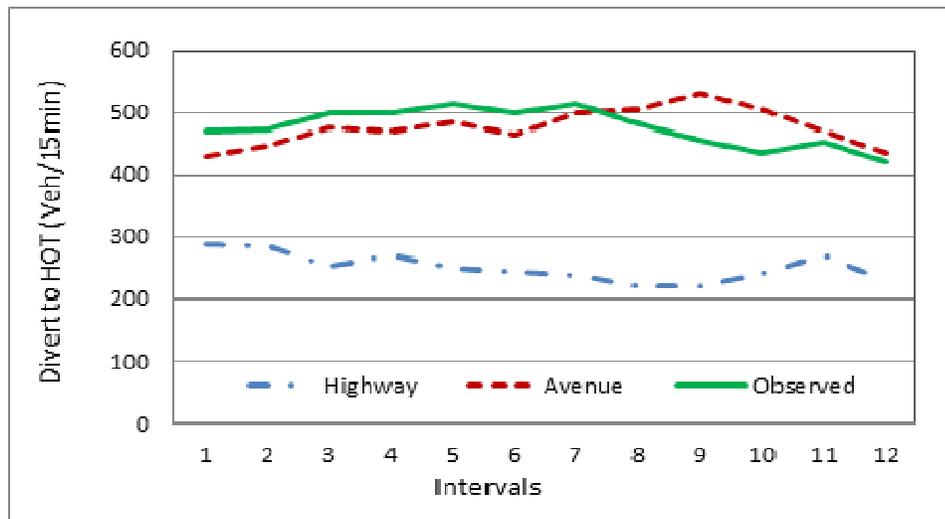


Figure 7-100 Comparison Between Modeled and Observed ML Volume for Willingness-to-Pay Curve Method

7.6. Convergence and Stability

Figure 7-111 compares the relative gap in the willingness-to-pay approach and in the generalized cost function approach. The willingness-to-pay approach shows very poor convergence. Figure 7-122 shows the diverted volume to the ML in each iteration for both assignment approaches. These figures clearly demonstrate the instability of the willingness-to-pay approach, as it applied to DTA in this study.

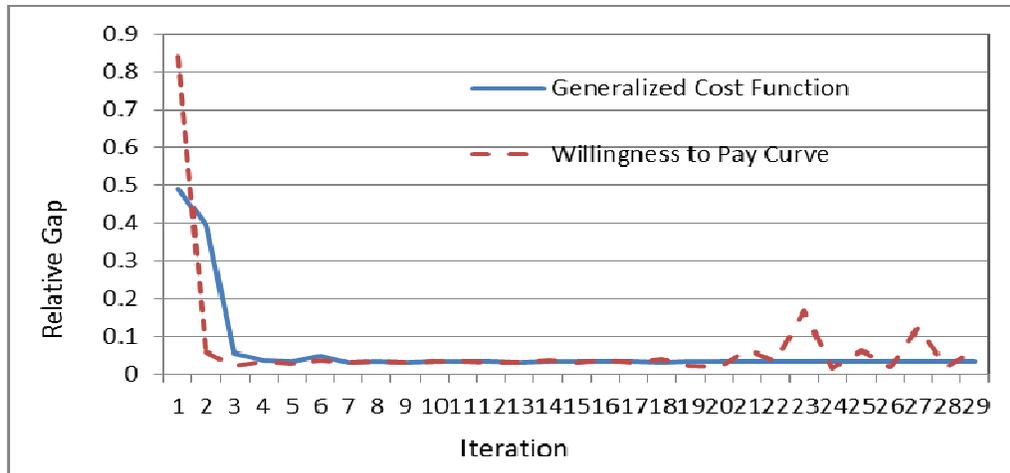


Figure 7-111 Relative Gap for Different Assignment Approaches

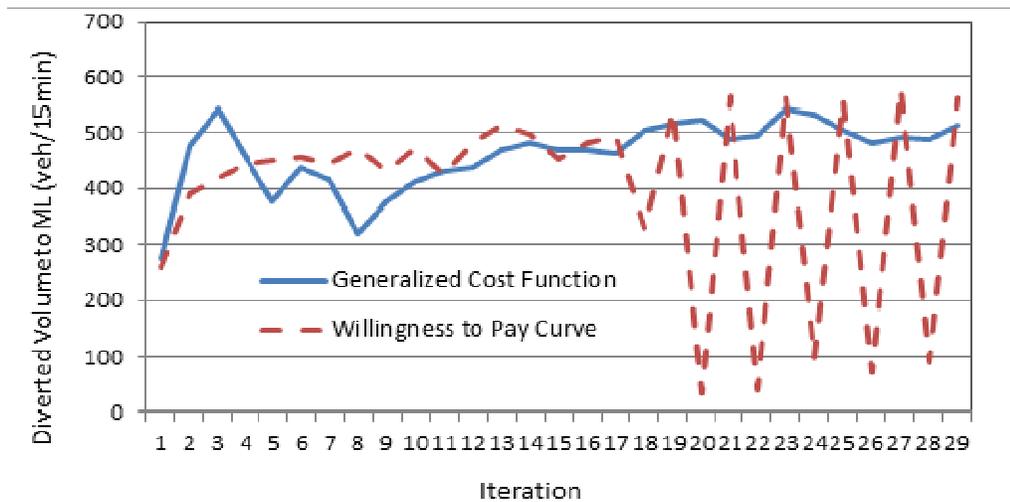


Figure 7-122 Diverted Volume to ML for Different Assignment Approaches

It should be noted that in general, and particularly in the case of managed lane modeling, trip-based or route-based measures of convergence are more important to be checked, compared to link-based convergence measures. The current versions of Highway and Avenue assignment modules report link-based convergence measures and do not report trip-based measures. Additional criteria such as the variations in the volume utilizing the ML in sequential iterations should be used for checking convergence. The

calibration procedure adapted in this study is an iterative process between demand, network and assignment parameters. The final iteration is conducted when the assignment and route choice behavior is calibrated. Once the assignment parameters are calibrated, the OD estimation procedure needs to be run one more time. With the final estimated OD trips, the network parameters should be fine-tuned again to replicate real-world congestion patterns. This will complete the iterative supply-demand-assignment processes.

7.7. Model Validation and Sensitivity Analysis

The demand and route choice parameters in this research were calibrated based on the volume averaged over representative days. The median day or any other day could have been used. In most locations, the volumes vary with a coefficient of variance (variance/ mean) of 3% to 7% between days. For validation and sensitivity analyses, Cube Avenue was run with different demand values from low to high, to see if it can replicate days with lower/higher congestion.

Figure 7-133 and Figure 7-144 show the speed contours for GPL, resulting from the generalized cost function and willingness-to-pay approaches, respectively. Demand is slightly changing in each scenario. It can be seen that the generalized cost function assignment can reasonably respond to the change in the demand level in terms of increased congestion patterns, meaning that the higher demands produce more congested networks. Unexpectedly, for the willingness to pay approach, demands that are 96 percent of the original demands produced a high congestion. The reason is that the assignment solution is not stable, as was shown in Figure 7-111 and Figure 7-122.

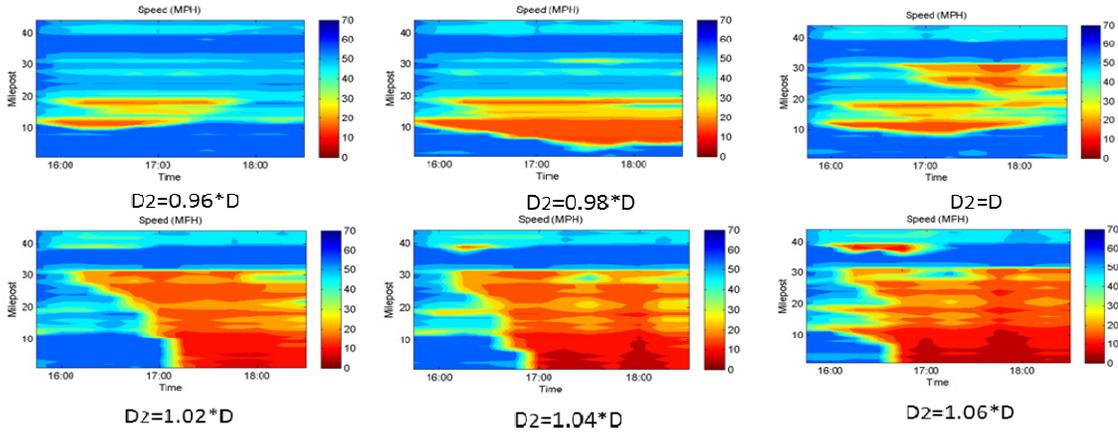


Figure 7-133 Speed Contour for GPL with Different Demand Level in Generalized Cost Function Assignment

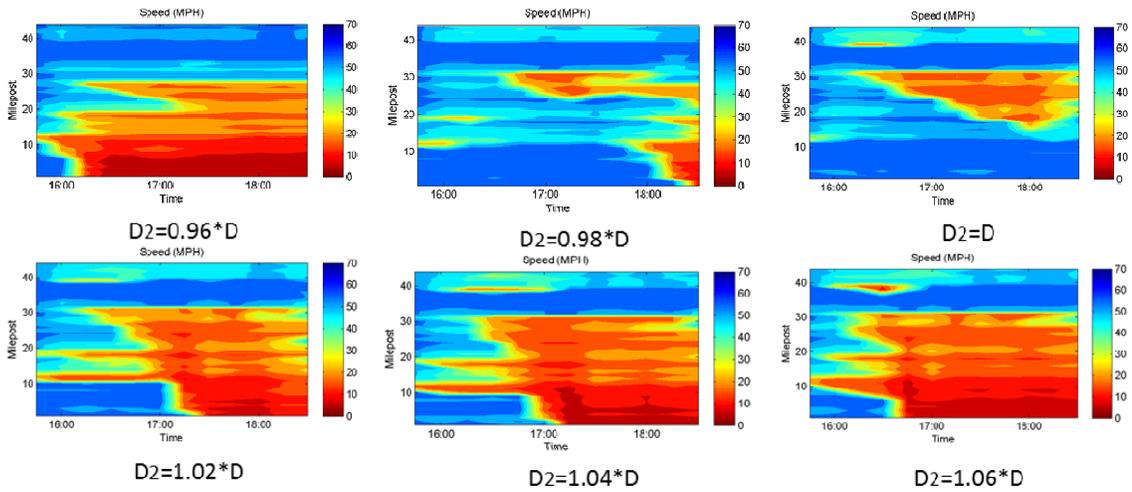


Figure 7-144 Speed Contour for GPL with Different Demand Level in Willingness-to Pay Curve Assignment

Figure 7-155 and Figure 7-166 show the trend of VMT and VHT with changing demands using the generalized cost approach and willingness-to-pay approach, respectively. As expected, in the stable generalized cost approach, VMT and VHT increase with increase in demand, but change randomly in the unstable willingness to pay approach.

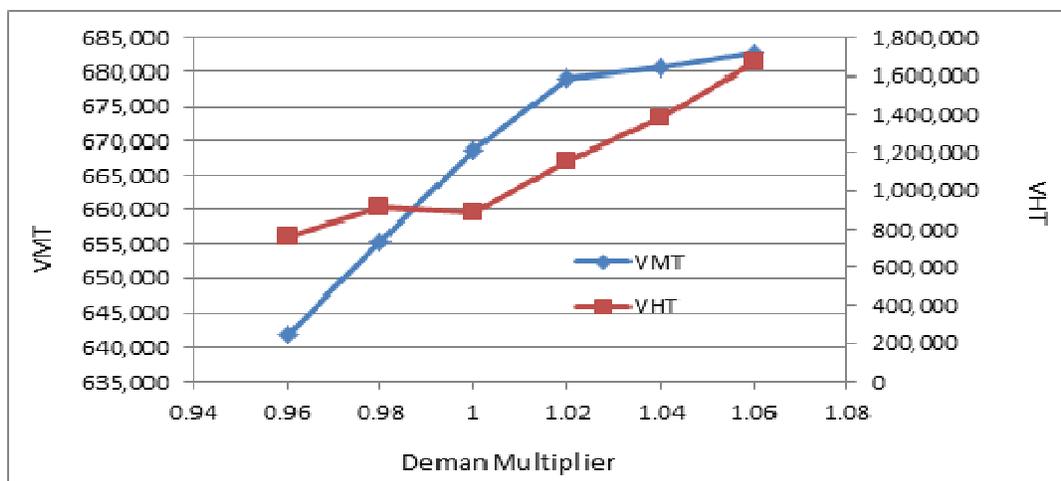


Figure 7-155 Changing in VMT and VHT with Changing Demand in Generalized Cost Function Assignment

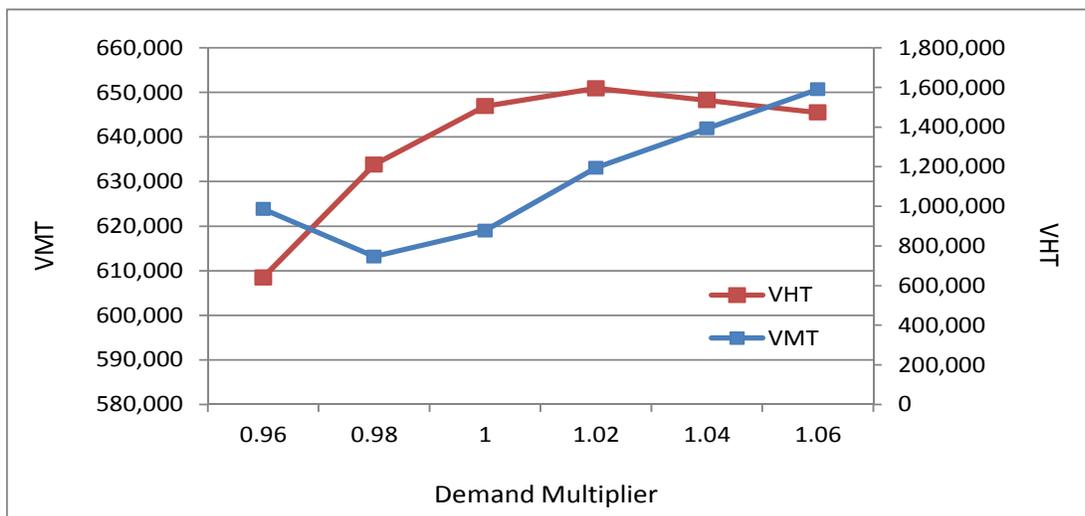


Figure 7-166. Changing in VMT and VHT with Changing Demand in Willingness-to-Pay Curve Assignment

Figure 7-177 demonstrates the effect of using different seed numbers in the simulation-based DTA tool. As can be seen, although it is tested on the converged, stable generalized cost function assignment, the results of the two runs are not identical. Thus, more research is needed to define how many runs are enough to account for the effect of randomness in the simulation.

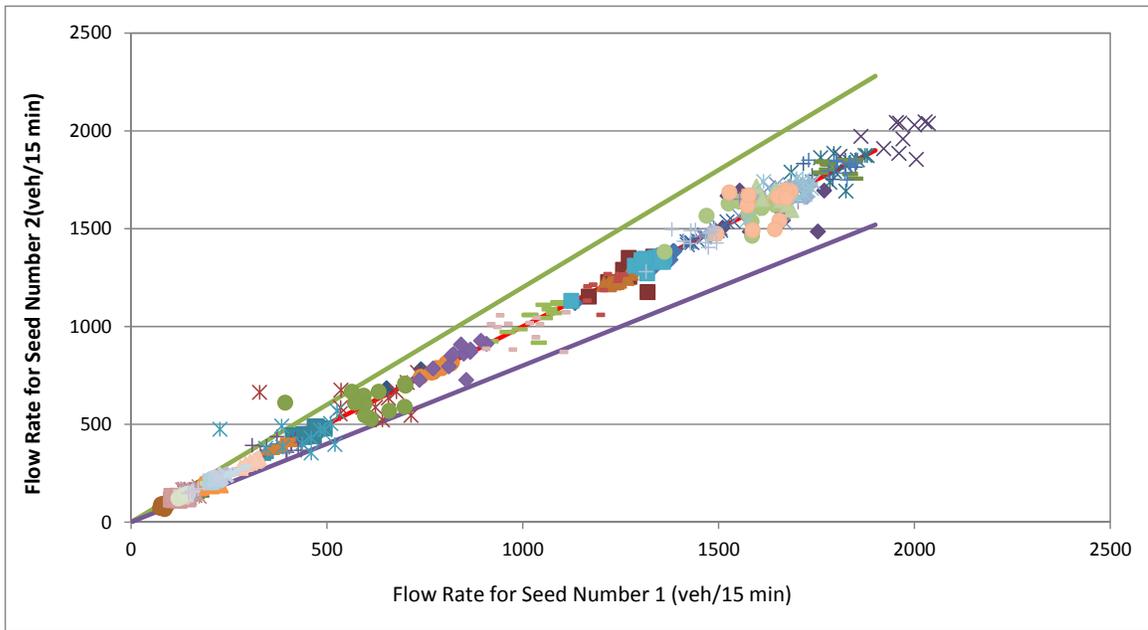


Figure 7-177. Effect of Randomness in Simulation-Based DTA Tool

7.8. Assessment of Performance Measures

During the calibration, extensive use of data visualization was conducted and volumes, speeds, and queues were compared to real-world measures by different goodness of fit measures. A number of state and FDOT standards were consulted in this process including:

- FSUTMS-Cube Framework Phase II Model Calibration and Validation Standards (Cambridge Systematics, Inc., 2008).
- Ohio RMSE Curve, which offers a target percent root mean squared error by volume group (Cambridge Systematics, Inc., 2010).

The final results for a converged and calibrated generalized cost function assignment are presented below. Figure 7-188 shows the scatter plot of observed versus simulated volumes at screenline locations. The coefficient of determination (R^2) between

the simulated and observed data is high (0.9761), indicating very high correlation with about 5% overestimation of the volumes on average as indicated by the 1.0515 coefficient value. It is interesting to compare this figure, with Figure 7-19, which is scatter plot for the same screenlines, for two different representative days. This figure demonstrates the day-to-day variation in real-world volume. The coefficient of determination between the volume measurements for these two days (0.7715), is lower than the one between observed and simulated values.

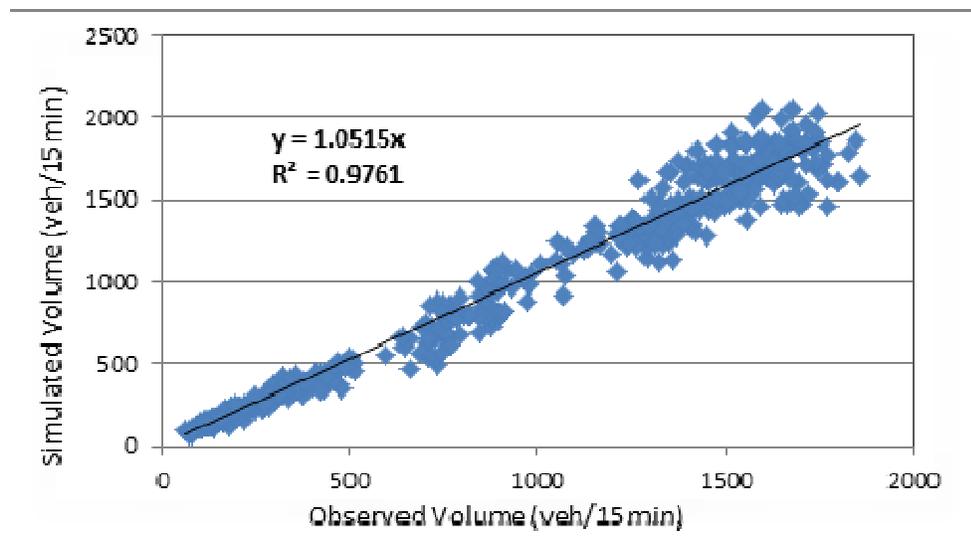


Figure 7-188 Scatter Plot of Observed vs. Simulated Volume of Screenlines

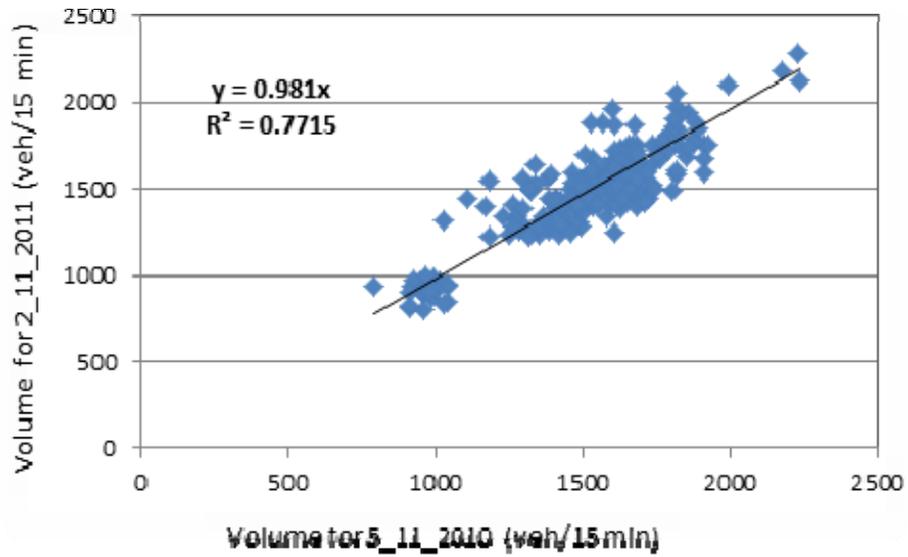


Figure 7-19 Scatter Plot of Observed Volume of Screenlines for Different Days

Figure 7-190 compares the observed and simulated flow rate on ML, which is an indicator of good calibration of route choice behavior.

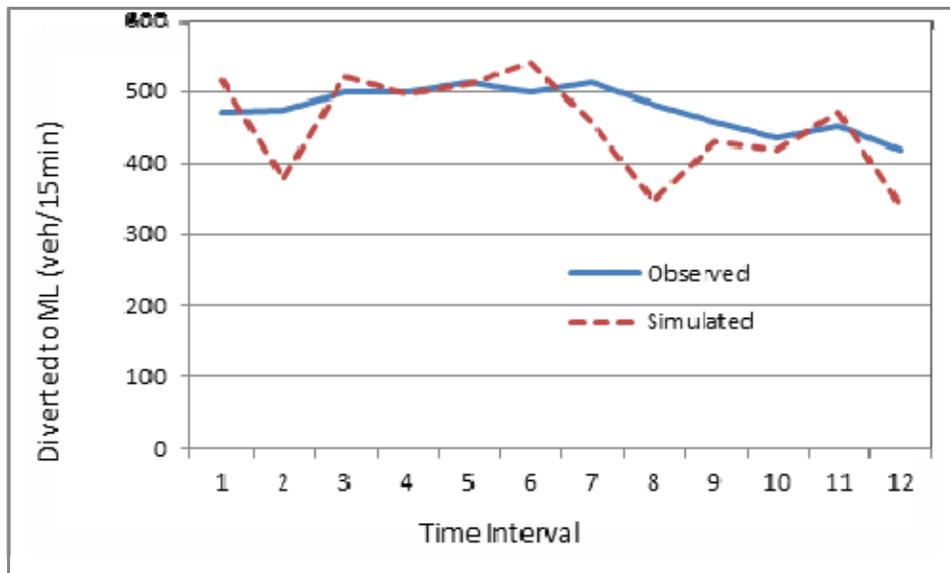


Figure 7-190 Comparison Between Observed and Simulated Flow on ML

Table 7-5 presents calculated goodness of fit measures of volume replication. The results in this table indicate an acceptable goodness of fit between the measured and simulated volumes.

Table 7-5 Goodness of Fit Statistics for Volume Replication

Goodness of Fit Statistics	Value
RMSE	113
% RMSE	13.6
MAE	77.0
R squared	0.976
GEH <5	87%
GEH <10	100%

Drawing RMSE curve for different ranges of volume showed that the resulting RMSE curve is well below the Ohio RMSE curve, which again indicate an acceptable volume estimation.

Figure 7-201 shows speed contour for simulated and one observed representative day, indicating a successful replication of real-world congestion pattern.

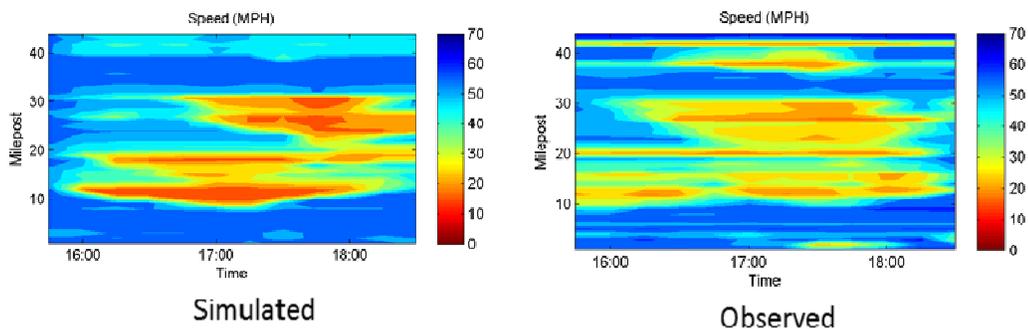


Figure 7-201 Comparison Between Simulated and Observed Speed Contour

Figure 7-212 is scatter plot of the simulated speed, versus one observed representative day. This figure indicates not a good correlation between the measured and simulated speeds, although Figure 7-201 shows that the model was able to model the

queues relatively well. Again it is interesting to compare the results in Figure 7-212 with those in Figure 7-223, which shows the relationship between the speeds for two different representative days. As can be seen, due to the probabilistic nature of traffic breakdown and traffic demands, there is a great variation in day-to-day congestion patterns and the correlation of speeds between these days is also low.

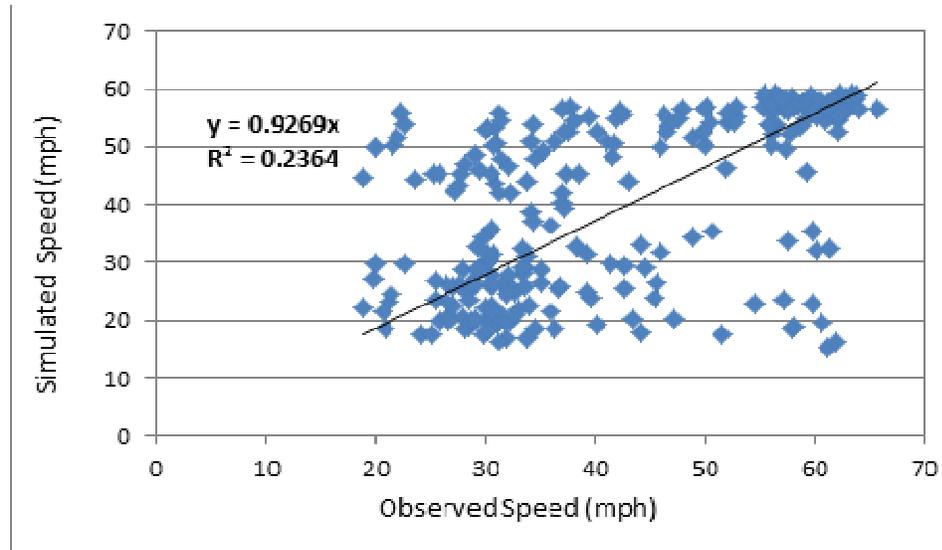


Figure 7-212 Scatter Plot of Observed vs. Simulated Speed of Screenlines

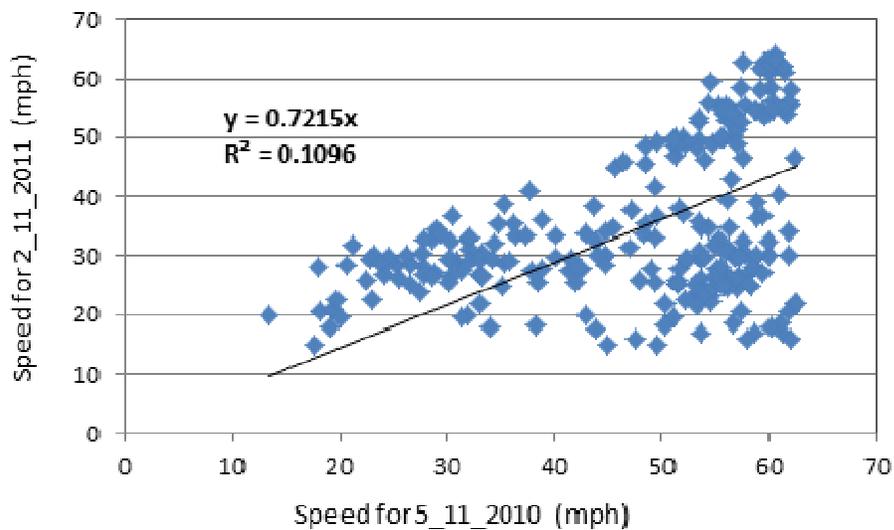


Figure 7-223 Scatter Plot of Observed Volume of Screenlines for Different Days

7.9. Summary

This chapter discussed two different approaches for modeling route choice behavior and illustrated the calibration procedure for each method for a real-world corridor. It was shown how toll schedule data and microwave detector data can be integrated and used in calibration. For the “generalized cost function” approach, the value of time is the most important parameter to calibrate. In this study, this parameter stands for the aggregated perceived benefit of using ML for the paid toll cost. This value was estimated to be \$42 for the case study corridor. The shape of the “willingness to pay” curve is the key parameter in the second approach. The initial curve was taken from another corridor, in another state, and it was shown in this study how dramatically the calibrated curve is deviated from the initial curve. This interestingly highlights the importance of site specific data and calibration processes.

It was found that although both approaches produce similar results in terms of predicting the percentage of travelers that use ML, the generalized cost function is more straightforward to implement and calibrate, and it also converges better.

The superiority of DTA over STA to replicate the percentage of ML users and to replicate the real-world congestion pattern was clearly illustrated. The quality of assignment calibration significantly affects the result of OD estimation and, consequently, the result of the network calibration. Incorrect estimation of assignment parameters may result in inaccurate ODs, and if the network is loaded with this trip table, the network and route choice parameter estimation would be adversely affected as well. It

is essential to re-run the OD estimation process and network calibration after the assignment is calibrated, and the whole procedure shown in

Figure 3-1 should be repeated until convergence is achieved.

8. CONCLUSION AND FUTURE STUDY

Simulation-based dynamic traffic assignment (DTA) has been increasingly utilized to evaluate traffic management strategies, including managed lanes (ML). Compared to traditional methods that normally utilize static traffic assignment (STA) and simple analytical traffic flow equations, simulation-based DTA better captures the dynamics of traffic operations by modeling time-variant system measures (including queuing and travel times), demand, advanced management strategies, and the associated responses of travelers. Although DTA has the potential to better replicate real-world conditions, the quality of the result is highly dependent on the quality and resolution of the input data and the adequacy of the calibration process. The present study shows how data from multiple sources can be integrated, validated, and best used in different stages of ML assignment modeling and calibration. Extensively and carefully processing demand, traffic, and toll data, as well as properly defining performance measures, resulted in a calibrated and stable model, which closely replicates real-world congestion patterns and can reasonably respond to perturbations in network and demand properties.

The following lessons were learned as a result of the present study:

- Advanced modeling tools such as DTA require more detailed and higher quality data to ensure that the developed model accurately replicates real-world conditions. This study successfully utilized detector data collected by the regional traffic management center, combined with PTMS ramp counts, toll data from traffic management center, and measurements from other sources of data to satisfy the DTA data needs. However, significant efforts

were required to process, fuse, and validate the data for use in the modeling processes.

- A sequential procedure that iterates between network calibration, demand estimation, and route choice parameter estimation is recommended in this study. Despite the existence of mathematical formulas and solutions for the simultaneous estimation of supply and demand parameters, their implementations in the real world are not straightforward and have not been executed properly.
- Supply or network calibration in Cube assignment tools entails estimating capacity, free-flow speed, and traffic flow model parameters for each link in the network. These parameters affect the travel time, queue formation, and queue spillback when the demand is loaded onto the network. A systematic multilevel approach to network calibration is recommended in this study, with an increasing calibration scope at each level. The process starts at the level of separated bottlenecks, where the capacity is estimated by various methods based on field data. The network is gradually extended to connected bottlenecks, and then to the whole corridor and subarea coverage. The advantage of this approach is twofold: First, critical spots of the network can be better identified, analyzed and replicated. Second, a more reliable demand can be estimated for the smaller network. Focusing on isolated bottleneck locations and the freeway corridor for managed lane assessment enables the capturing of the interactions between supply and demand and makes it feasible to track the causes for congestion.

- The supply calibration performed in this study illustrates the importance of coding the capacity based on detector measurements in DTA tools, particularly when there is evidence that the capacity is lower than the HCM-based estimates. In the case explored in this study, it was found that the free-flow speed and more importantly, the capacity, were overestimated by the HCM procedure, resulting in incorrect travel times and congestion when used in the DTA model.
- One of the important congestion spots in the modeled network is caused by spillback from an off-ramp that causes low speeds in the two left lanes of the five-lane road segment. Since the utilized DTA tool does not support lane-by-lane modeling, it was not possible to correctly replicate that location, because the queue in the model first fills up the whole segment (including five lanes) before backing up to the upstream link. If replicating the congestion at such locations is important to a study, a tool that better handles this situation or multi-resolution analysis should be considered.
- During the matrix estimation process with the currently available tools, several manual adjustments and iterations are required to ensure joint calibration of demand, supply, and route choice behaviors. Automated methods such as the Method of Successive Average (MSA) or a state-space framework may reduce the need for manual adjustment. However, as long as there is evidence that a specific OD pair needs to be changed, manual adjustment remains a necessary component of OD estimation. Adjustments

and fine-tunings are also needed to avoid unrealistic deviation from the initial matrix and trip pattern.

- When calibrating supply, demand, and assignment parameters, a distance function between simulation outputs and field measurements is minimized. This function should include different measures, such as link volumes, OD demands, link speeds and/or densities, etc. Limiting the function to replicating link volumes, as is the case in many studies, can be misleading and fail to produce the correct demand or congestion pattern. Most OD matrix estimation methods are based on link traffic volumes and initial OD matrices. If enough data on speeds, densities, queue lengths, OD routes or zonal trips (production and attraction rates) are available, they should be incorporated into the calibration process to better replicate real-world traffic conditions. This study illustrates that for a congested network, if queue or density data is not incorporated in the OD estimation process, the resulting trip matrix can replicate real-world counts but fail to produce the congestion pattern, because it can only replicate the capacity constrained counts, and not the real demand.
- Dynamic traffic assignment requires trip matrices specified for short time intervals (e.g., 15 minutes or 30 minutes). The derivation of these matrices is performed in this study using a sequential process that starts from matrix factorization based on count data, followed by static assignment-based OD matrix estimation (static OD estimation), and finally followed by dynamic assignment-based OD matrix estimation (dynamic OD estimation). However, identified limitations, tool immaturity, and the results of this study indicate

that in its current stage, the dynamic OD estimation process in the utilized tool should be used with caution until further enhancements and testing of these enhancements are completed so as to confirm that the tool is able to produce reliable results.

- Calibrating the toll curve, value of time, and willingness-to-pay curve parameters are important aspects of ML assignment. The quality of the result in this step significantly affects demand estimation and network calibration. A willingness-to-pay curve with an uncalibrated shape or incorrect value of time can distort the result of network calibration and demand estimation. Therefore, feedbacks from the assignment step to network calibration and demand estimation are necessary.
- There is evidence that the value of time used in the SERPM model (\$12.60 per hour) is low and that a value of \$42 produces better results.
- The findings from this study highlight the shortcomings of utilizing static assignment for assessing managed lanes, even when the measured capacity values are coded, illustrating the need to utilize DTA modeling for such assessments. The calibrated DTA model was able to produce results that are similar to real-world results. However, the Cube static assignment module was not able to replicate real-world conditions.
- For the case study of this research, it was found that the “generalized cost function” approach and the “willingness-to-pay” approach produce comparable results, although the generalized cost approach is much simpler to implement and calibrate, and can converge to a stable solution.

There are several limitations in the present study that need to be addressed in future research. Time restraints also limited the scope of this study. Some potentially interesting topics to explore in future studies include:

- exploring a dynamic OD estimation process in which DTA is used to assign the matrix in the lower-level of the bi-level optimization framework,
- estimating driver population factors, based on socioeconomic data, and their impact on capacity estimation for the case study corridor,
- using a logit model instead of a willingness-to-pay curve and calibrating it based on observed counts on ML,
- comparing the result of the aforementioned logit model calibration with the result of the logit model based on a local stated preference survey,
- stratifying the value of time and willingness-to-pay curves for different markets. This classification can be based on user income, trip purpose, vehicle occupancy, or a combination of these variables.

REFERENCES

- Akcelik, R. (2003). Speed-Flow and Bunching Relationship for Uninterrupted Flows Presented at 25th Conference of Australian Institute of Transport Research, University of South Australia.
- Alvarez, P. (2012). A Methodology to Estimate Time Varying User Responses to Travel Time and Travel Time Reliability in a Road Pricing Environment. Ph.D. Dissertation, Florida International University
- Antoniou, C., Ben-Akiva, M., and Koutsopoulos, H. (2007). Non-linear Kalman Filtering Algorithms for On-line Calibration of Dynamic Traffic Assignment Models. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 8, No. 4, pp. 661 –670.
- Ashok, K. and Ben-Akiva, M. (1993). Dynamic OD Matrix Estimation and Prediction for Real-Time Traffic Management Systems. In Daganzo, C., editor, *Transportation and Traffic Theory*, Elsevier Science Publishing, pp. 465–484.
- Ashok, K., and Ben-Akiva, M. (2002). Estimation and Prediction of Time-Dependent Origin–Destination Flows with a Stochastic Mapping to Path Flows and Link Flows, *Transportation Science.*, Vol. 36, No. 2, pp. 184–198.
- Balakrishna, R. (2002). Calibration of the Demand Simulator in Dynamic Traffic Assignment System, Master Thesis, Massachusetts Institute of Technology.
- Balakrishna, R., Ben-Akiva, M., and Koutsopoulos, H. (2007). Off-line Calibration of Dynamic Traffic Assignment: Simultaneous Demand-Supply Estimation. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2003, pp. 50-58.
- Bar-Gera, H., Boyce, D. (1999). Route Flow Entropy Maximization in Origin-Based Traffic Assignment. In Ceder, A. (Ed.), *In Proceedings of the 14th International Symposium on Transportation and Traffic Theory*, Jerusalem, Israel, 1999. Elsevier Science, Oxford, UK, pp. 397–415.
- Bar-Gera, H. (2010). Traffic Assignment by Paired Alternative Segments. *Transportation Research Part B*, Vol. 44, pp.1022-1046.
- Bassan, S. and Polus, A. (2010). Meaning of Actual Capacity of Freeways and Its Estimation. *Canadian Journal of Civil Engineering*, Vol. 37, No. 1, pp. 77-87.

- Bertini, R. L., Fernandez, R., Wieczorek, J., Li, H. (2008). Using Archived ITS Data to Automatically Identify Freeway Bottlenecks in Portland, Oregon. In: Proceedings of the 15th World Congress on ITS, New York.
- Bertini, R.L. and Myton, A. (2005). Using PeMS Data to Empirically Diagnose Freeway Bottleneck Locations in Orange County, California. Transportation Research Record: Journal of the Transportation Research Board, No. 1925, pp. 48-57.
- Boyce, D., Nie, Y., Bar-Gera, H., Liu, Y., and Hu, Y. (2010). Field Test of a Method for Finding Consistent Route Flows and Multiple-Class Link Flows in Road Traffic Assignments. Final Report to the Federal Highway Administration, Washington, DC.
- Boyce, D., O'Neill, C. R., and Scherr, W. (2008). Solving the Sequential Travel Forecasting Procedure with Feedback. Transportation Research Record: Journal of the Transportation Research Board, No. 2077, pp. 129-135.
- Boyce, D., Ralevic-Dekic, B. and Bar-Gera, H. (2002). Convergence of Traffic Assignments: How much is Enough? Journal of Transportation Engineering, Vol. 130, No. 1, pp. 49-55.
- Brandiss, J. (2001). Estimation of Origin-Destination Flows for Dynamic Traffic Assignment. Master Thesis, Massachusetts Institute of Technology.
- Brilon, W., Geistefeldt, J., and Regler, M. (2005). Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity, In: Proceedings of the 16th Symposium on Transportation and Traffic Theory.
- Burris, M., Nelson, S., Kelly, P., Gupta, P. and Cho, Y. (2012). Willingness-to-pay for HOT Lanes - Empirical Analysis from I-15 and I-394. Transportation Research Record: Journal of the Transportation Research Board, No. 2297, pp. 47-55.
- Cambridge Systematics, Inc. (2008). FSUTMS-Cube Framework Phase II Model Calibration and Validation Standards. Prepared for Florida Department of Transportation Systems Planning Office, http://www.fsutmsonline.net/images/uploads/reports/FR2_FDOT_Model_CalVal_Standards_Final_Report_10.2.08.pdf, Accessed Oct 12, 2013.
- Cambridge Systematics, Inc. (2010). Travel Model Validation and Reasonableness Checking Manual (Second Edition). Prepared for Federal Highway Administration, <http://www.camsys.com/pubs/FHWA-HEP-10-042.pdf>, Accessed Oct. 12, 2013.
- Cambridge Systematics, Inc. (2012). Value of Travel Time Reliability, Synthesis Report and Workshop Working Paper. Prepared for SHRP 2 Workshop on Value of Travel Time Reliability.

<http://onlinepubs.trb.org/onlinepubs/shrp2/L35RFP/DraftSynthesis-ValuingReliability.pdf>, Accessed Oct 12, 2013.

Chao, C., Varaiya, P., and Kwon, J. (2005). An Empirical Assessment of Traffic Operations. In Proceedings of the International Symposium on Transportation and Traffic Theory.

Chen, C., Skabardonis, A., and Varaiya, P. (2004). Systematic Identification of Freeway Bottleneck, In: Proceedings of the 83rd Transportation Research Board Annual Meeting, Washington, D.C.

Chen, Z. (2003). Bayesian filtering: From Kalman filters to particle filters, and beyond. Journal of Statistics, Vol. 182, No.1, pp: 1-69.

Chi, H., (2010). An Improved Framework for Dynamic Origin-Destination Matrix Estimation. PhD. Dissertation, Florida International University.

Chiu, Y.C. (2012). Modeling Express Lanes Using Dynamic Traffic Assignment Models. FDOT Managed Lane Workshop. May 22, 2013, http://www.fsutmsonline.net/images/uploads/12_Dynamic_Traffic_Modeling_-_Yi-Chang_Chiu.pdf, Accessed Oct. 15, 2013.

Chiu, Y. and Bustillos, B. (2009). A Gap Function Vehicle Based Solution Procedure for Consistent and Robust Simulation Based Dynamic Traffic Assignment. 88th Annual Meeting of the Transportation Research Board.

Chiu, Y., Bottom, J., Mahut, M., Paz, A., Balakrishna, R., Waller, T. and Hicks, J. (2011). Dynamic traffic assignment: A primer. Transportation Research E-Circular, No. E-C153.

Chiu, Y. C., Zhou, L., and Song, H. B. (2010). Development and Calibration of the Anisotropic Mesoscopic Simulation model for Uninterrupted Flow Facilities. Transportation Research, Part B, No. 44 pp.152-174.

Citilabs, Inc. (2013). Citilabs Cube User Manual.

Davidson, K.B. (1966). A Flow–Travel Time Relationship for Use in Transportation Planning. In Proceedings of the 3rd ARRB Conference, Vol. 3, No. 1, pp. 183-194.

Davidson, K.B. (1978). The Theoretical Basis of a Flow–Travel Time Relationship for Use in Transportation Planning. Australian Road Research, Vol. 8, No. 1, pp. 32-35.

Dehghani, Y., Adler, T., Doherty, M. W., and Fox, R. (2003). Development of a new toll mode-choice modeling system for Florida's Turnpike Enterprise. Transportation Research Record: Journal of the Transportation Research Board, No. 1858, pp. 9-17.

Dehman, A. (2012). Breakdown Maturity Phenomenon at Wisconsin Freeway Bottlenecks. In: Proceedings of the 91st Annual Meeting of Transportation Research Board.

Dervisoglu, G., Gomes, G., Kwon, J., and Varaiya, P. (2009). Automatic Calibration of the Fundamental Diagram and Empirical Observations on Capacity, In Proceedings of the 88th Annual Meeting of Transportation Research Board.

Dixon, M. P. and Rilett, L. R. (2002). Real-time OD estimation using automatic vehicle identification and traffic count data. *Journal of Computer-Aided Civil and Infrastructure Engineering*, Vol. 17, No.1, pp. 7-21.

Doan D.L., Ziliaskopoulos, A. and Mahmassani, H.S. (1999). An On-line Monitoring System for Real-Time Traffic Management Applications. *Transportation Research Record: Journal of the Transportation Research Board*, No.1678, pp. 142-149.

Doblas, J. and Benitez, F. (2005). An Approach to Estimating and Updating Origin–Destination Matrices Based Upon Traffic Counts Preserving the Prior Structure of a Survey Matrix. *Transportation Research Part B*, Vol. 39, No. 7, pp. 565–591.

Dowling, R., Skabardonis, A., and Alexiadis, V. (2004). *Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Software*. Final Report, FHWA-HRT-04-040.

Dowling, R. G., Kittelson, W., Skabardonis, A., and Zegeer, J. (1997). *Techniques for Estimating Speed and Service Volumes for Planning Applications*. NCHRP Report 387: TRB, National Research Council, Washington, D.C.

Elefteriadou, L., Roess, R.P., and McShane, W.R. (1995). Probabilistic Nature of Breakdown at Freeway Merge Junctions. *Transportation Research Record*, No.1484.

Elefteriadou, L. and Lertworawanich, P., (2003). Defining, Measuring, and Estimating Freeway Capacity. Presented at the 82nd Annual Meeting of the Transportation Research Board.

Federal Highway Administration (FHWA), (2008a). Policy Options Evaluation Tool for Managed Lanes (POET-ML) Users Guide and Methodology Description, Federal Highway Administration HOV Lane Performance. <http://ops.fhwa.dot.gov/publications/fhwahop09031/>, Accessed Oct. 10, 2013.

Federal Highway Administration (FHWA), (2008b). Value Pricing Pilot Program TRUCE 3.0 Users Guide, Federal Highway Administration HOV Lane Performance. http://ops.fhwa.dot.gov/tolling_pricing/value_pricing/tools/truce_model_guide.htm, Accessed Oct. 10, 2013.

- Federal Highway Administration (FHWA), (2008c). Managed Lanes: A Primer. http://ops.fhwa.dot.gov/publications/managelanes_primer/managed_lanes_primer.pdf, Accessed Oct. 20, 2013.
- Florida Department of Transportation (FDOT), (2013). Southeast Florida Regional Planning Model, v. 6.5.3 (SERPM). Florida Department of Transportation, Tallahassee, Florida. http://www.fsutmsonline.net/index.php?/model_pages/model_pages/, Accessed Oct. 20, 2013.
- Florida's Turnpike Enterprise, (2005) 2005 Annual Customer Service Satisfaction Survey, http://www.floridasturnpike.com/downloads/CustomerSurveys/2005custsat_web.pdf, Accessed Nov 4th, 2013.
- Goodall, N., and Smith, B. L. (2010). What Drives Decisions of Single-Occupant Travelers in High-Occupancy Vehicle Lanes? Transportation Research Record: Journal of the Transportation Research Board, No. 2178, 156-161.
- Gupta, A. (2005). Observability of Origin-Destination Matrices for Dynamic Traffic Assignment, Master's thesis, Mass. Inst. Technol
- Hadi, M, Ozen, H, Shabaniyan, S, Xiao, Y, Zhao, and W, Ducca, F. (2012). Use of Dynamic Traffic Assignment in FSUTMS in Support of Transportation Planning in Florida. Final Report Submitted to the Florida Department of Transportation.
- Hall, F. L., and Agyemang-Duah, K. (1991). Freeway Capacity Drop and the Definition of Capacity, Transportation Research Record: Journal of the Transportation Research Board, No. 1320, pp. 91-98.
- Hansen, S., Byrd, A., Delcambre, A., Rodriguez, A., and Bertini, R.L. (2005). Using Archived ITS Data to Improve Regional Performance Measurement and Travel Demand Forecasting 2005 CITE Quad/WCTA Regional Conference Vancouver, BC, Canada.
- Hu, S. R. and Chen, C. B. (2004). Dynamic Estimation of Freeway Origin-Destination Demand and Travel Time Using Extended Kalman Filtering Algorithm, In Proceedings of IEEE Conference on Networking, Sensing and Control, pp.1329-1334.
- Huntsinger, L., and Roupail, N. (2011). Calibrating Travel Demand Model Volume-Delay Function Using Bottleneck and Queue Analysis. Presented at 90th Annual Meeting of the Transportation Research Board, Washington DC.
- Janson, M. and Levinson, D. (2013). HOT or Not, Driver Elasticity to Price on the MnPASS HOT Lanes. <http://nexus.umn.edu/Papers/HOTorNOT.pdf>, Accessed Oct. 30, 2013.

- Jia, A., Williams, B.M., and Rouphail, N.M. (2010). Identification and calibration of site specific stochastic freeway breakdown and queue discharge. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2188, pp. 148-155.
- Kunde, K. K. (2002). Calibration of Mesoscopic Traffic Simulation Models for Dynamic Traffic Assignment. Master Thesis. Massachusetts Institute of Technology.
- Lin, W. (2006). A Robust Model for Estimating Freeway Dynamic Origin-Destination Matrices. PhD. Dissertation, Department of Civil and Environmental Engineering University of Maryland, College Park. MD.
- Lin, D.Y., (2009). A Dual Approximation Framework For Dynamic Network Analysis: Congestion Pricing, Traffic Assignment Calibration, and Network Design Problem, Ph.D. Dissertation, University of Texas at Austin.
- Lu, S., and Nie, Y., (2010). Stability of User-equilibrium Route Flow Solutions for the Traffic Assignment Problem. *Transportation Research Part B*, Vol. 44, No. 4, pp.609–617.
- Loudon, W. R. (2007). Improving the Estimation of Potential Travel-Time Savings from HOV Lanes. In *Transportation Research Board Annual Meeting*. Washington, DC.
- Mahmassani, H., Qin, X., and Zhou, X. (2004). DYNASMART-X Evaluation for Real-Time TMC Application: Irvine Test Bed. Technical report, Maryland Transportation Initiative, University of Maryland.
- Minderhoud, M.M., Botma, H., and Bovy, P.H.L. (1997). Roadway Capacity Estimation Methods Explained and Assessed. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1572.
- National Cooperative Highway Research Program (NCHRP), (2012). Assessing Highway Tolling and Pricing Options and Impacts, Volume II: Travel Demand Forecasting Tools. Transportation Research Board, Washington, D.C. http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_722v2.pdf, Accessed Oct., 20, 2013.
- Nava, E., Milster, K., Chiu, Y.C., (2013). Modeling Managed Lane/Express Lane With Dynamic Traffic Assignment, Submitted for presentation at the 93rd Transportation Research Board Annual Meeting, Washington D.C.
- Nguyen, S. (1982). Estimating Origin-Destination Matrices from Observed Volumes. In *Proceedings of the First Course on Transportation Planning Models of the International School of Transportation Planning*, Amalfi, Italy.

Ortúzar, J.D. and Willumsen, L.G. (2001). Modeling Transport (3rd Edition). John Wiley and Sons, Chichester, England.

Parsons Brinckerhoff, (2013). Managed Lane Modeling Application for FSUTMS, Phase II, Memorandum to Advanced Traffic Assignment Subcommittee. Florida Department of Transportation,
http://www.fsutmsonline.net/images/uploads/ManagedLane_Phase_II_Draft_2_2.docx,
Accessed Oct. 23, 2013.

Peeta, S., Kumar, A., and Sharma, S. (2011). Post-Processing Techniques to Enhance Reliability of Assignment Algorithm Based Performance Measures. Publication FHWA/IN/JTRP-2011/19. Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana, doi: 10.5703/1288284314643.

Rakha, H., and Arafeh, M. (2010). Calibrating Steady-State Traffic Stream and Car-Following Models Using Loop Detector Data. Transportation Science, Vol. 44, No. 2, pp. 151-168.

Resource Systems Group, (2012). South Florida Expressway Stated Preference Survey Report. Prepared for URS Corporation.

Resource Systems Group, (2010). Convergence Peer Exchange Read Ahead Information.

Ruegg, S., Kulshrestha, A., and Puppala, S. (2013). Phase I Managed Lane Assignment Model Testing I-95 Corridor, Orlando and Tampa Regional Model. Memorandum to Advanced Traffic Assignment Subcommittee, Florida Department of Transportation,
http://www.fsutmsonline.net/images/uploads/ManagedLaneTesting_PhaseI.docx,
Accessed Oct. 20, 2013.

Saberi, M. (2010). Investigating Freeway Speed-Flow Relationships for Traffic Assignment Applications M.S. Thesis, Portland State University.

Safirova, E., Gillingham, K., Harrington, W., and Nelson, P. (2003). Resources for the Future, Are HOT Lanes a Hot Deal?
<http://www.rff.org/rff/Documents/RFF-IB-03-03.pdf>, Accessed Oct. 20, 2013.

Schrank, D. L. and Lomax, T. J. (2007). The 2007 Urban Mobility Report, Texas Transportation Institute, Texas A & M University.

Singh, R. (1995). Beyond the BPR Curve: Updating Speed-Flow and Speed-Capacity Relationships in Traffic Assignment. Presented at 5th Conference on Transportation Planning Methods Applications, Seattle, Washington.

- Sloboden, J., Lewis, J., Alexiadis, V., Chiu, Y., and Nava, E. (2012). Traffic Analysis Toolbox Volume XIV: Guidebook on the Utilization of Dynamic Traffic Assignment in Modeling. Final Report, FHWA-HOP-13-015, U.S. Department of Transportation, Federal Highway Administration.
- Spiess, H. (1990a). Conical Volume-Delay Functions. *Transportation Science*, Vol. 24, No. 2.
- Spiess, H. (1990b). A Gradient Approach for the OD Matrix Adjustment Problem. Vol. 693. Publication.
- Tavana, H. (2001). Internally-consistent Estimation of Dynamic Network Origin-Destination Flows from Intelligent Transportation Systems Data Using Bi-level Optimization. Ph.D. Dissertation, University of Texas at Austin.
- Transportation Research Board (2000). Highway Capacity Manual, Washington, D.C.
- Transportation Research Board (2010). Highway Capacity Manual, Washington, D.C.
- Tung, R. and Chiu, Y.C. (2011). Integration of Dynamic Traffic Assignment in a Four-Step Model Framework, A Deployment Case Study in PSRC (Puget Sound Regional Council) Model. Transportation Research Board Planning Applications Conference, Reno, Nevada.
- URS, Inc. (2011). Toll Road Modeling Techniques, Published by the Maricopa Association of Governments, Headquartered in Phoenix, AZ. <http://www.azmag.gov/addons/MAG/download.asp?ID=10890>, Accessed Nov. 3, 2013.
- Van Aerde, M., and Rakha, H. (1995). Multivariate Calibration of Single Regime Speed-Flow-Density Relationships VNIS/Pacific Rim Conference, Seattle, WA.
- Van Arem, B. and van der Vlist, M. J. M. (1992). An On-line Procedure for Estimating Current Capacity. Technical Report INRO-VVG 1991-17, TNO Institute of Spatial Organization, Delft.
- Vaze, V. S. (2007). Calibration of Dynamic Traffic Assignment Models with Point-to-Point Traffic Surveillance. Master thesis. Massachusetts Institute of Technology.
- Vovsha, P., Davidson, W., and Donnelly, R. (2006). Making the State of the Art the State of the Practice: Advanced Modeling Techniques for Road Pricing. In Expert Forum on Road Pricing and Travel Demand Modeling.
- Vovsha, P., Donnelly, B., and Bradley, M. (2013). Impact of Congestion Pricing and Travel Time Reliability on Travel Demand, Managed Lane Modeling Practice Workshop. Miami, Florida.

- Wang, Y. et al. (2009). An adaptive freeway traffic state estimator, *Automatica*, Vol. 45, pp.10-24.
- Wang, Y., Liu, X., and Roupail, N. (2012). Analysis of Managed Lanes on Freeway Facilities, Final Report for NCHRP Project 03-96.
- Wang, Y., Papageorgiou, M., Messmer, A., Coppola, P., Tzimitsi, A., and Nuzzolo. A. (2009). An Adaptive Freeway Traffic State Estimator. *Automatica*. Vol. 45, pp.10-24.
- Washburn, Scott S., Modi, V., and Kulshrestha, A. (2010). Investigation of Freeway Capacity: a) Effect of Auxiliary Lanes on Freeway Segment Volume Throughput and b) Freeway Segment Capacity Estimation for Florida Freeways. No. TRC-FDOT-73157-2010.
- WilburSmith Associates, (2011). Investment Grade Traffic and Revenue Study US. 36 Managed Lane. Final Report submitted to Colorado Department of transportation, http://www.coloradodot.info/library/studies/us-36-managed-lanes-investment-grade-traffic-and-revenue-study/Final-20Report_080911.pdf, Accessed Oct 25, 2013.
- Xiao, Y., Hadi, M., Ozen, H., and Mysore, V. (2010). An ITS Evaluation Tool in the FSUTMS Regional Demand Modeling Environment. Transportation Research Record: Journal of the Transportation Research Board, No. 2176, pp. 76-83.
- Yeon, J., Hernandez, S., and Elefteriadou, L., (2007). Differences in Freeway Capacity by Day of the Week, Time of Day, and Segment Type. In Proceedings of the 86th Annual Meeting of the Transportation Research Board, Washington, D.C.
- Zhang, L. and Levinson, D. (2004). Some Properties of Flows at Freeway Bottlenecks. In: Proceedings of the 83rd Transportation Research Board Annual Meeting, Washington, D.C.
- Zhou, X. (2004). Dynamic Origin-Destination Demand Estimation and Prediction for Offline and Online Dynamic Traffic Assignment Operation. Ph.D. Dissertation, University of Maryland.
- Zijpp, N. J. and Hamerslag, R. (1994). Improved Kalman Filtering Approach for Estimating Origin-Destination Matrices for Freeway Corridors. Transportation Research Record: Journal of the Transportation Research Board, No. 1443, pp. 54-64.

VITA

SHAGHAYEGH SHABANIAN

EDUCATION

- 2001 – 2006 B.S., Civil Engineering
Iran University of Science and Technology, Tehran, Iran
- 2008 – 2009 M.S., Transportation Engineering
Sharif University of Technology, Tehran, Iran
- 2009 – 2014 Doctoral Candidate in Civil Engineering
Florida International University, Miami, Florida

PUBLICATIONS AND PRESENTATIONS

- Hadi, M., Ozen, H., and Shabanian, S., “Processes to Guide the Application of Dynamic Traffic Assignment by Transportation Agencies,” Proceedings of the 91st Annual Meeting of the Transportation Research Board, National Research Council, Washington, D.C., January 2012.
- Hadi, M., Ozen, H., Shabanian, S., Xiao, Y., Zhao, W., and Ducca, F., “Use of Dynamic Traffic Assignment in FSUTMS in Support of Transportation Planning in Florida” Final Report Submitted to the Florida Department of Transportation, June 2012.
- M. Hadi, Y. Xiao, P. Hu, A. Ramirex, and S. Shabanian, “Impact of Modeling Parameters on the HCM-Based Procedure to Estimate Reliability of Freeway Corridors.” Accepted for Publication at the Transportation Research Record: Journal of the Transportation Research Board, 2014.
- Shabanian, S. and Hadi, M., “Capacity Estimation in Support of Mesoscopic Simulation as Part of Dynamic Traffic Assignment Models,” presented at the 93rd Annual Meeting of the Transportation Research Board, National Research Council, Washington, D.C., January 2014.